Lesson 6

Analysis and Interpretation of Surveillance Data

Objectives for Lesson 6

- Describe methodological approaches to surveillance analysis
- Describe practical approaches to surveillance analysis
- Present surveillance data by time, place, and person
- Describe concept of rates and standardization of rates
- Describe approaches to exploratory data analysis
- Demonstrate uses of graphics and maps
- Describe systematic interpretation of surveillance data

Considerations in Analysis of Surveillance Data

- Know inherent idiosyncrasies of data set
- Proceed from simplest to most complex
- Realize when inaccuracies in data preclude more sophisticated analyses

Key Concepts that Determine Accuracy of Surveillance Data

Reliability:

Is a particular condition reported consistently by different observers?

Validity:

Does the particular condition as reported reflect the true condition as it occurs?

Basic Epidemiological Parameters

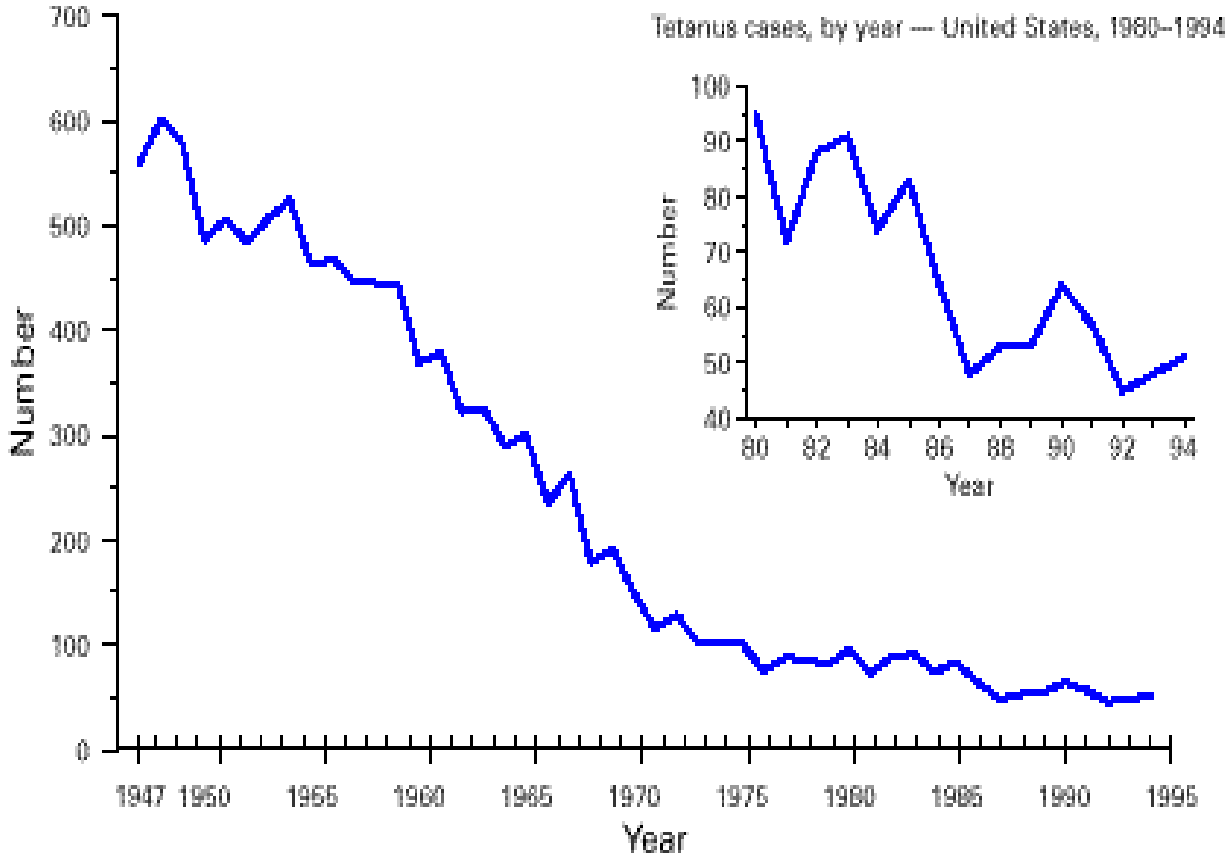
- Time
- Place
- Person

Data Analysis by Time

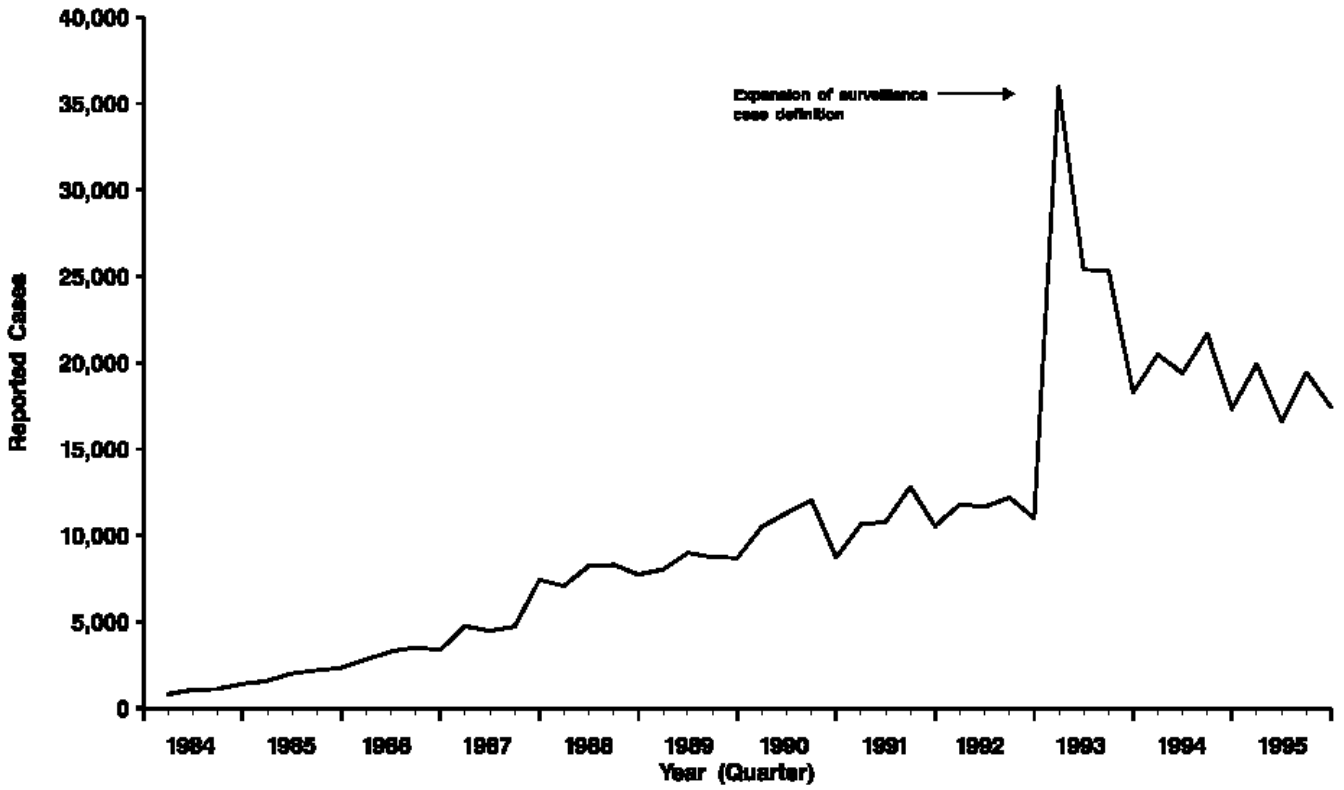
- Compare number of case reports received during a specific interval

- Compare number of cases for a current time period with number reported during same interval during previous time period

FIGURE 1. Reported number of tetanus cases, by year — United States, 1947–1994



ACQUIRED IMMUNODEFICIENCY SYNDROME (AIDS) — reported cases, by quarter, United States,* 1984–1995



*Includes Guam, Puerto Rico, the U.S. Pacific Islands, and the U.S. Virgin Islands.

The number of AIDS cases reported during 1995 was lower than the number reported in 1994 or in 1993. This decrease reflects the waning effect of the expansion, in 1993, of the AIDS case definition used for surveillance.

17

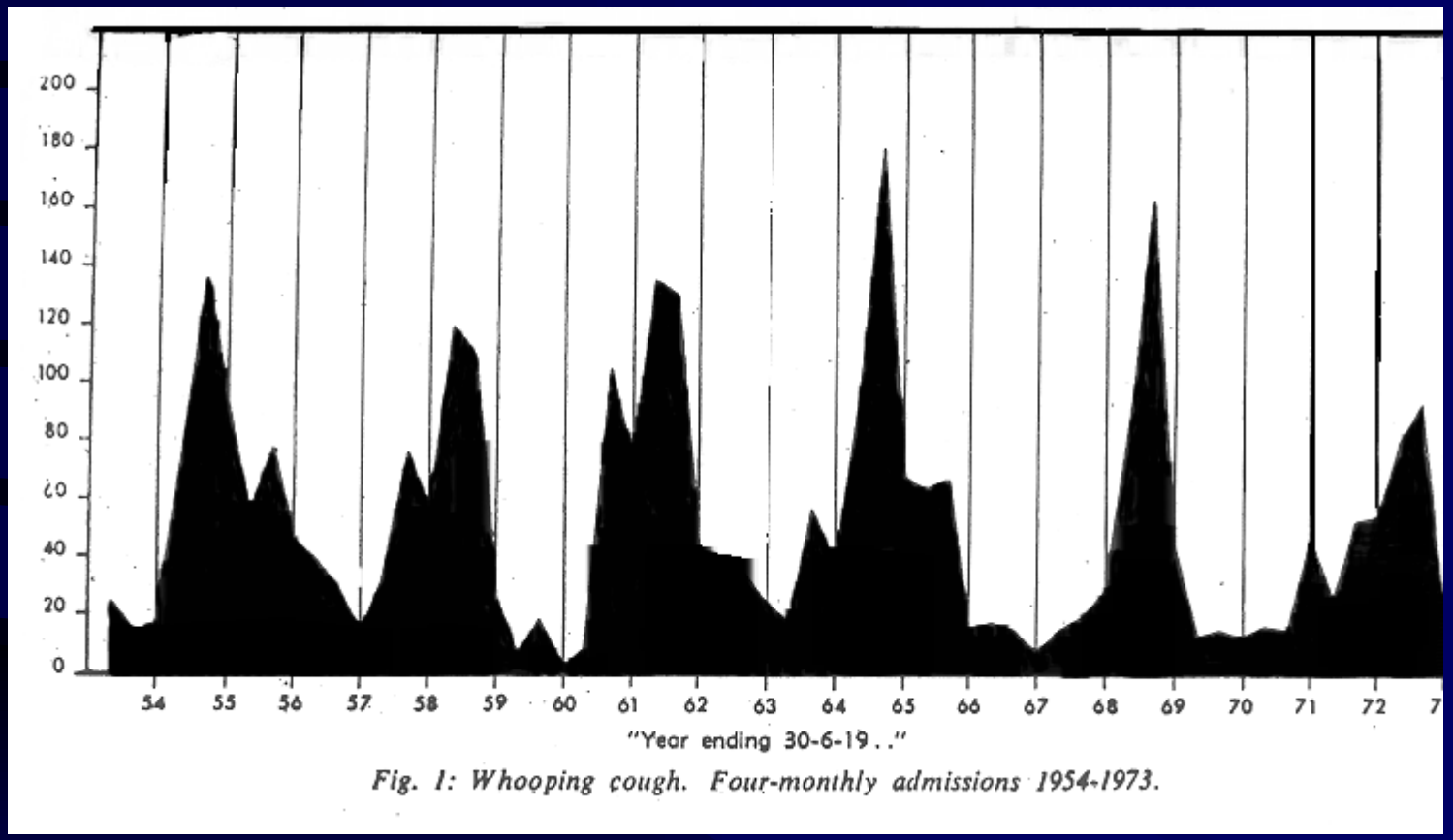


Fig. 1: Whooping cough. Four-monthly admissions 1954-1973.

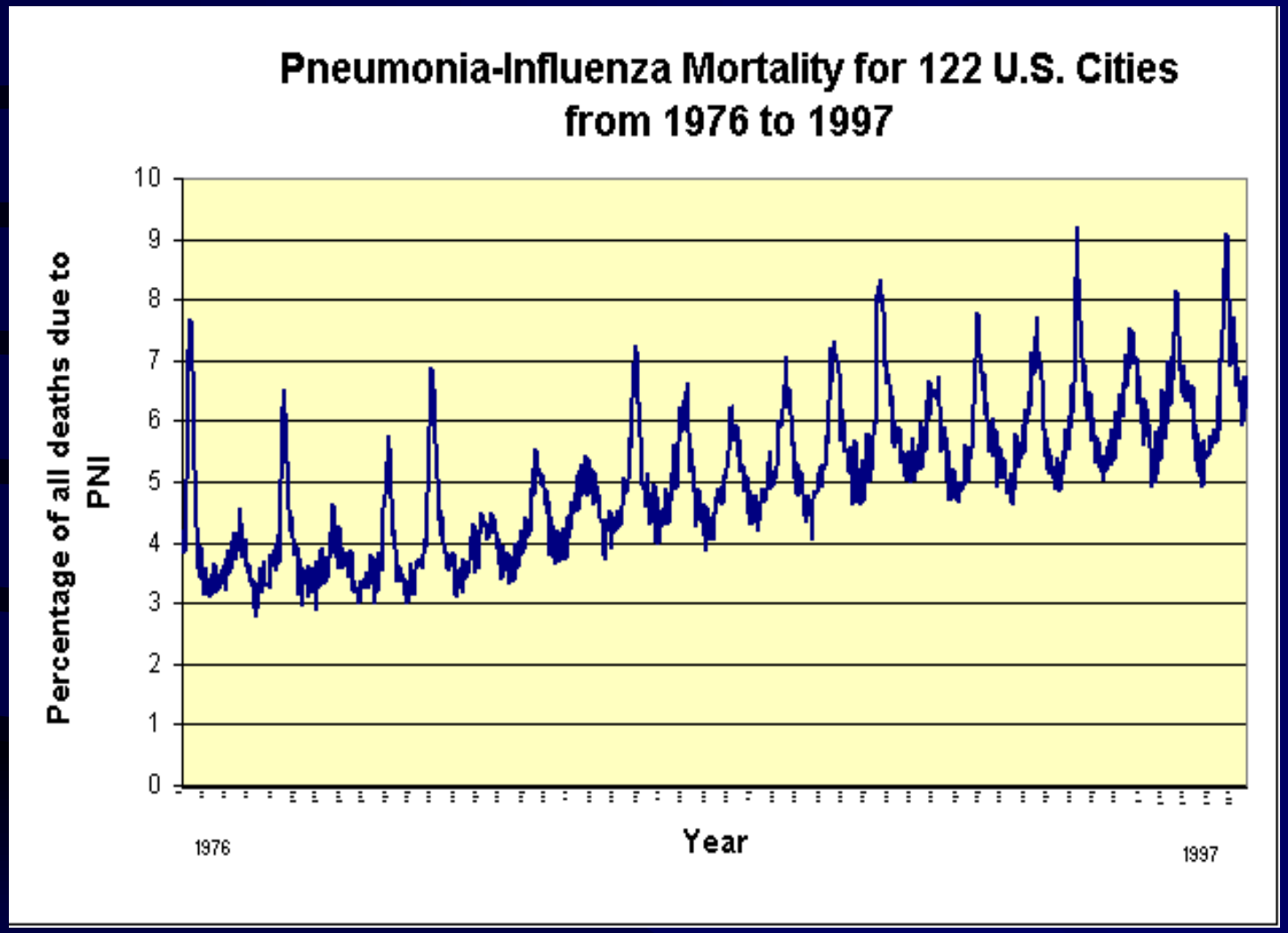
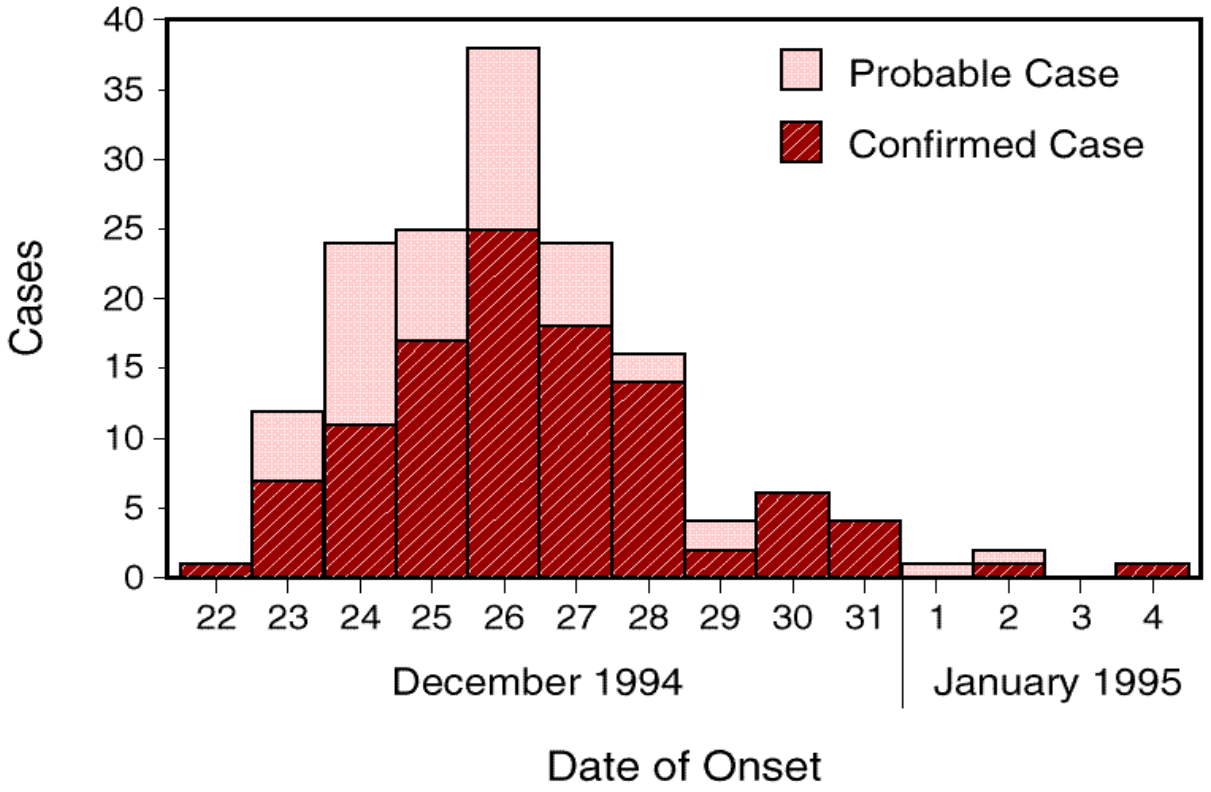

Pneumonia-Influenza Mortality for 122 U.S. Cities from 1976 to 1997
Percentage of all deaths due to PNI
10
9
8
7
6
5
4
3
2
1
0
1976
Year
1997

FIGURE 1. Number of probable* and confirmed† cases of *Salmonella* serotype Typhimurium infection, by date of onset — Wisconsin, December 22, 1994–January 4, 1995



*Diarrhea or abdominal cramps with onset during December 22, 1994–January 4, 1995, in a resident or a visitor to Dodge County or any of the four contiguous counties.
†Stool culture positive for tartrate-negative *Salmonella* Typhimurium.

Data Analysis by Place

- Where exposure occurred, not where it's reported from
- Allows prevention resources to be targeted effectively
- Use of computers and spatial mapping software allows for sophisticated analysis

BROAD STREET
PUMP
MARSHALL STREET
OUFOURS PLACE
BREWERY
NEW STREET
HUSBAND ST
LITTLE WINDMILL STREET
GREAT PULTENEY
BRIDLE
UP JAMES ST

Data Analysis by Person

Possible variables:

age

gender

race or ethnicity

marital status

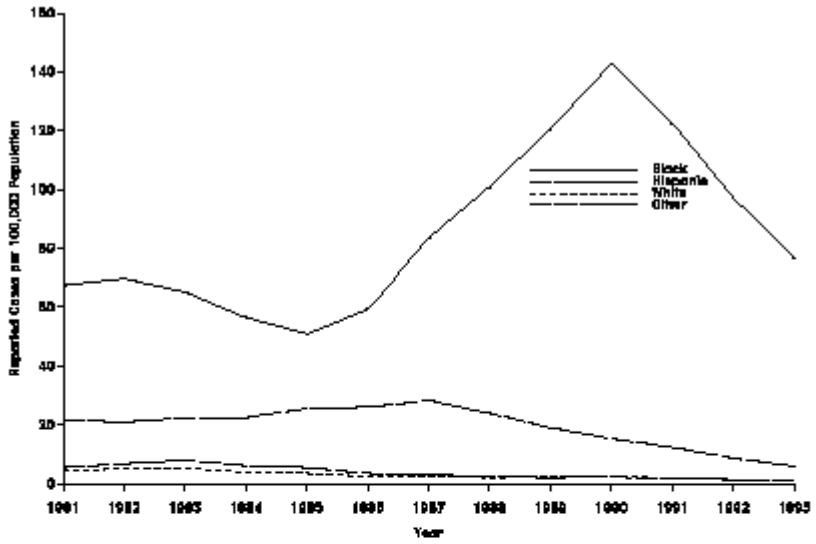
occupation

levels of income and education

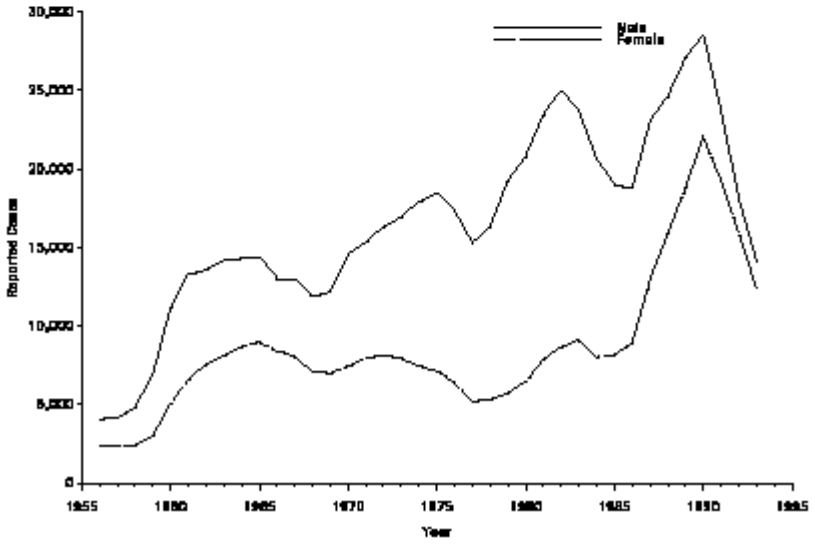
Interactions Among Time, Place, and Person

- Interactions can obscure patterns of disease and injury
- Syphilis in the 1980s

SYPHILIS (primary and secondary) — by race and ethnicity, United States, 1981–1993



SYPHILIS (primary and secondary) — by sex, United States, 1956–1993



Rate = Frequency of an Event

Numerator:

number of occurrences of an event during a specified time

Denominator:

size of population in which the event occurs

Ratio / Proportion / Rate

Ratio: any quotient obtained by dividing one quantity by another

Proportion: special type of ratio in which numerator is a subset of the denominator population

Rate: may be a proportion or may be limited in scope

Use of Rates in Epidemiology

- To formulate and test hypotheses about causes
- To identify risk factors for disease and injury
- To provide valid comparisons within or among populations for specific times

Types of Rates

- **Crude Rates:** describe a population
- **Specific Rates:** examine differences between a subpopulation and the entire population
- **Standardized Rates:** adjust rates when comparing populations

When should indirect standardization be used?

- when any of the specific rates in the study population are unavailable
- when such small numbers exist in categories of strata that data are unreliable

Data that must be available to use indirect adjustment to a rate:

- specific rates for the selected population
- distribution for the study population across the same strata as those used in calculating specific rates
- crude rate for the study population
- crude rate for the standard population

Advantages of Standardized Rates

- Adjust for confounding variables
- Provide measure which is easy to compare
- Have smaller standard error than specific rates
- Are more accurate and stable than specific rates
- Are more available for certain groups

Disadvantages of Standardized Rates

- May mask the difference
- Magnitude is arbitrary and depends on standard population

Analysis of Rate

1. Calculate crude rates
2. Compute relevant specific rates
3. Select standard population (if appropriate)
4. Present information using tables or maps
5. Apply statistical procedure

   (z test, Poisson parameters, chi-square analyses, time series methods)

Exploratory Data Analysis

- Serves as initial step in analysis
- Provides enumerative, graphic detective work
- Minimizes assumptions
- Allows data to motivate analysis
- Combines ease of description with quantitative knowledge

Steps in EDA

1. Use visual displays to convey structure of data and analyses
2. Transform data mathematically to simplify distribution
3. Investigate influence of outliers
4. Examine residuals

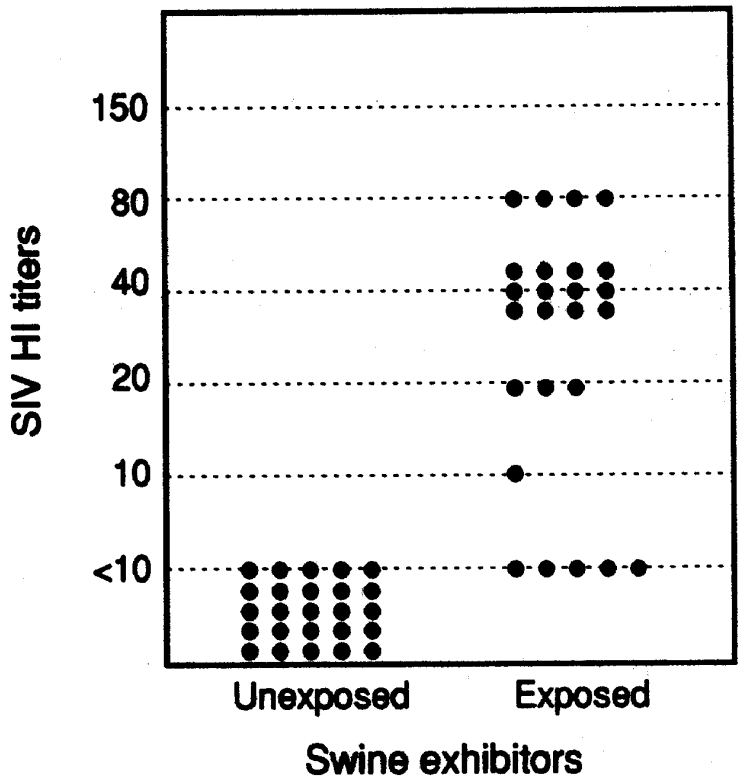


Figure 5–2 Dot plot of results of swine influenza virus (SIV) hemagglutination-inhibition (HI) antibody testing among exposed and unexposed swine exhibitors—Wisconsin, 1988.

ANALYZING AND INTERPRETING SURVEILLANCE DATA

1987: 226, 307, 350, 236, 222, 258, 197, 167, 138, 108, 191, 190, 201

1988: 216, 238, 331, 270, 265, 156, 164, 142, 112, 111, 153, 138, 159

1989: 145, 306, 314, 264, 222, 195, 155, 149, 102, 117, 174, 158, 159

Stem	Leaf
34	0
32	1
30	674
28	
26	450
24	8
22	22668
20	16
18	0157
16	474
14	259356899
12	88
10	28127

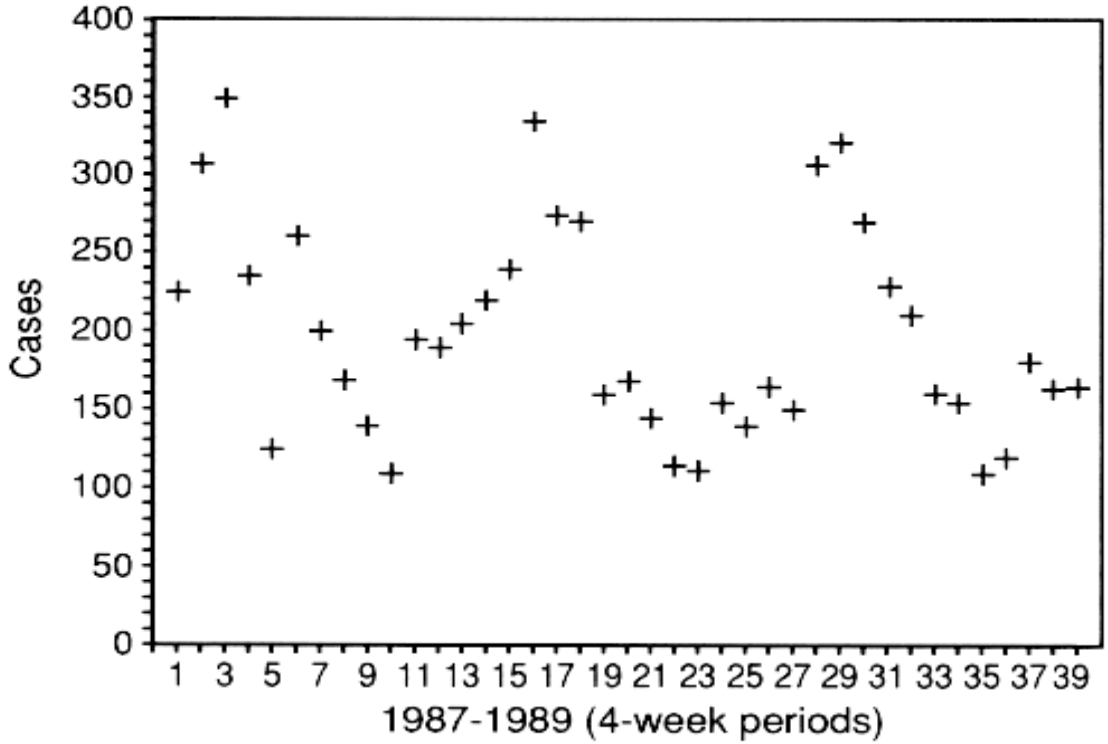


Figure 5–4 Scatter plot of 39 4-week totals of reported cases of meningococcal infections—United States, 1987–1989.

Table 5–6 Five-Number Summary of 39 4-Week Totals of Reported Cases of Meningococcal Infections—United States, 1987–1989

Median		190	
Hinges	151		237
Extremes	102		350

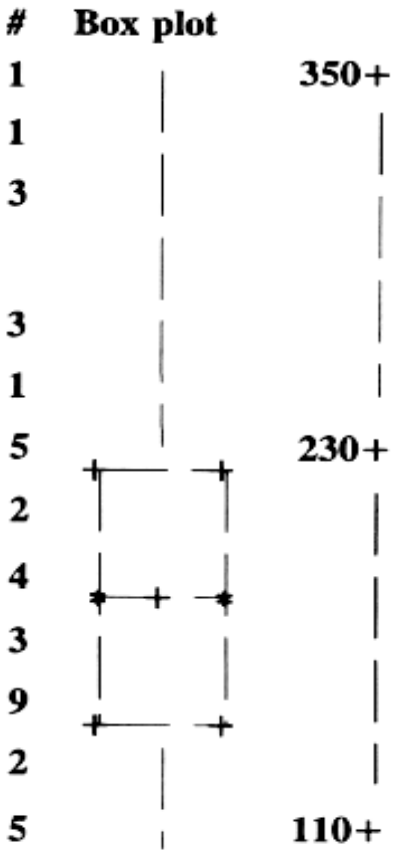


Figure 5–5 Box plot of 39 4-week totals of reported cases of meningococcal infections—United States, 1987–1989.

Purpose of Graphics

- To visually display measured quantities
- To allow researchers to mesh presentation and analysis
- To organize, summarize, and display information clearly and effectively

Tables

- Arrange data in rows and columns
- Demonstrate data patterns and relationships among variables
- Serve as a source of information for other types of data graphics

Guidelines for Developing a Table

- Describe what, when, where in the title
- Label rows and columns clearly
- Provide units of measure
- Provide row and column totals
- Define abbreviations and symbols
- Note data exclusions
- References Source
- Should stand alone

Table 5–9 Primary and Secondary Morbidity from Syphilis, by Age Category—United States, 1989

Age group (years)	*Cases*	
	Number	*Percentage**
≤14	230	0.5
15–19	4,378	10.0
20–24	10,405	23.6
25–29	9,610	21.8
30–34	8,648	19.6
35–44	6,901	15.7
45–54	2,631	6.0
≥55	1,278	2.9
Total	44,081	100.0

*Percentages do not add to 100.0 due to rounding.

TABLE 2. Gonorrhea rates* for 15- to 19-year-olds, by region, race/ethnicity, and sex — United States, 1991

Region	White		Black		Hispanic		Total population	
	Male	Female	Male	Female	Male	Female	Male	Female
Northeast†	41.8	136.5	7,061.6	7,325.6	720.7	749.3	516.3	701.1
South§	124.4	325.0	5,677.1	5,080.4	228.7	329.3	1,378.3	1,427.7
Midwest	82.4	264.8	6,012.6	5,790.8	74.7	119.0	897.6	1,149.6
West	61.3	209.0	3,956.5	3,569.6	230.7	231.6	365.0	485.3

*Per 100,000 population.
†Excludes cases from New York.
§Excludes cases from Kentucky and Maryland.

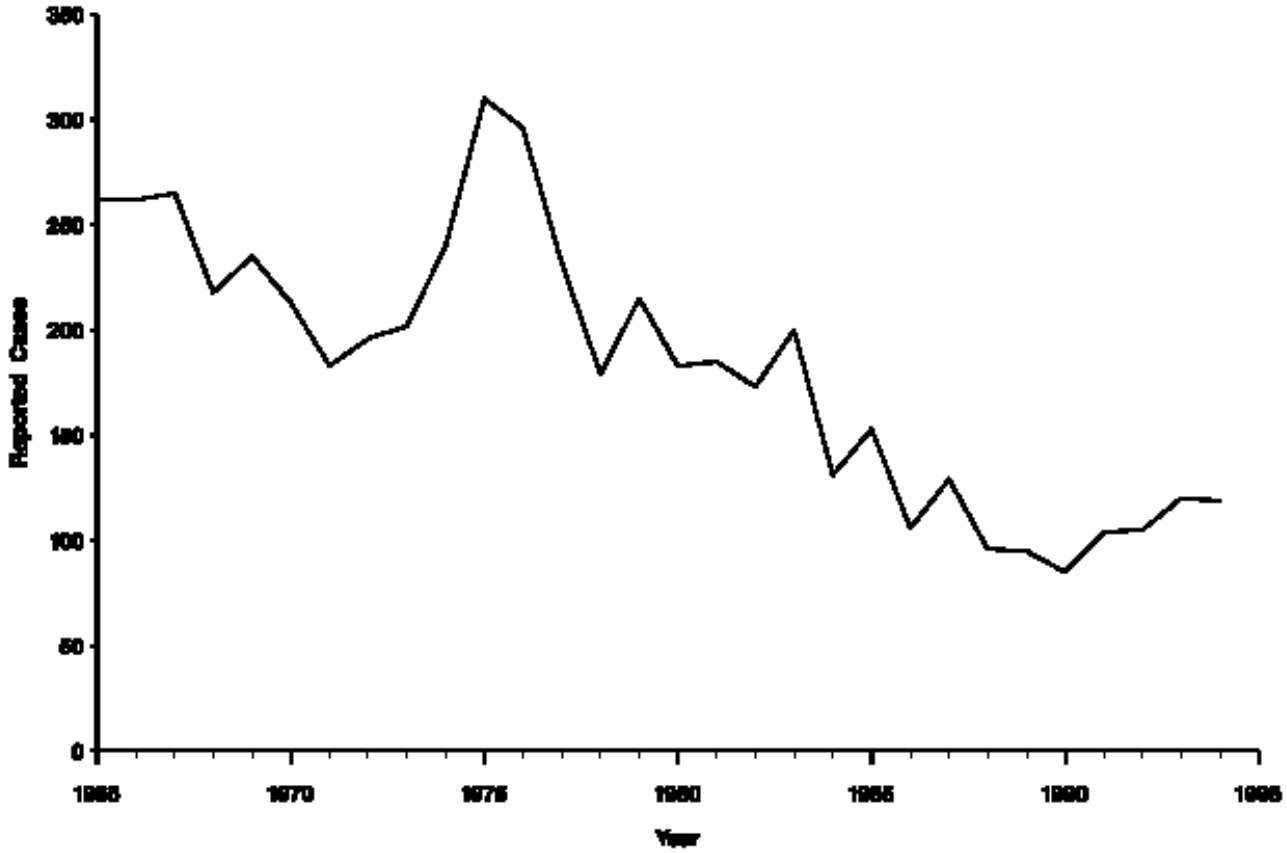
Graphs

- Visually display quantitative information
- Provide system of coordinates
- Assist reader to visualize patterns and trends

Guidelines in Developing Graphs

- Label title, source, axes, scales, legend
- Minimize the number of coordinate lines
- Portray frequency on the vertical scale, starting with zero
- Portray method of classification on the horizontal scale
- Indicate units of measure
- Define abbreviations and symbols
- Note data exclusions

BRUCELLOSIS — by year, United States, 1965–1994



•Recent data suggest that reported cases of foodborne brucellosis, usually caused by *Brucella melitensis*, are increasing in the United States.

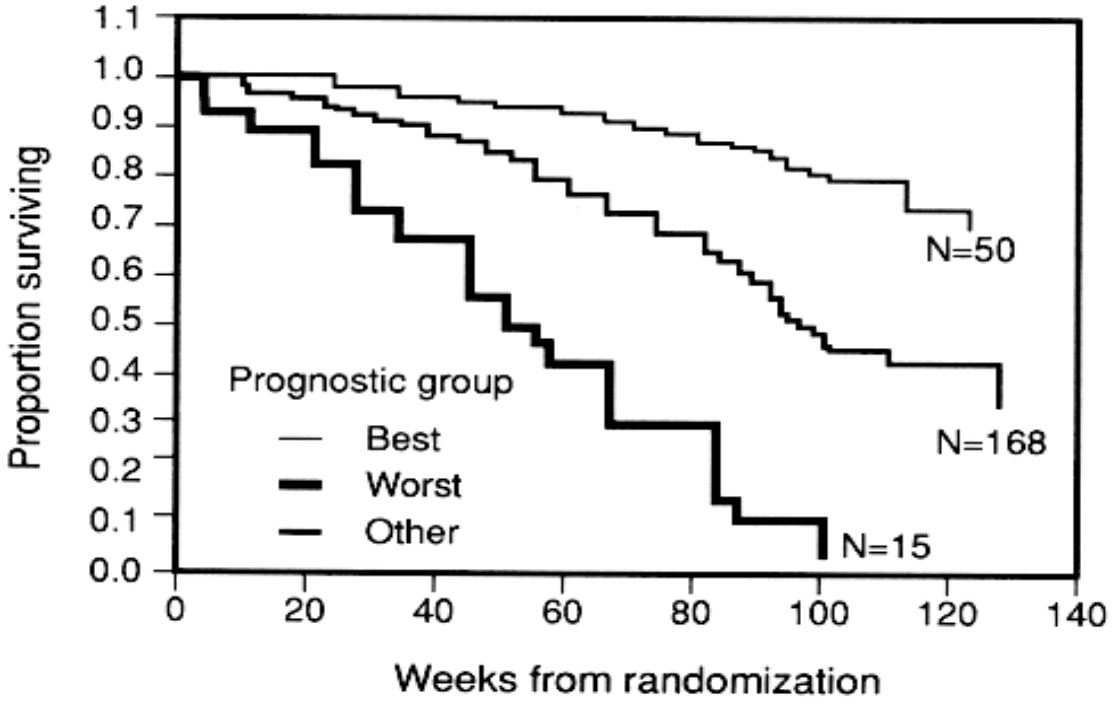
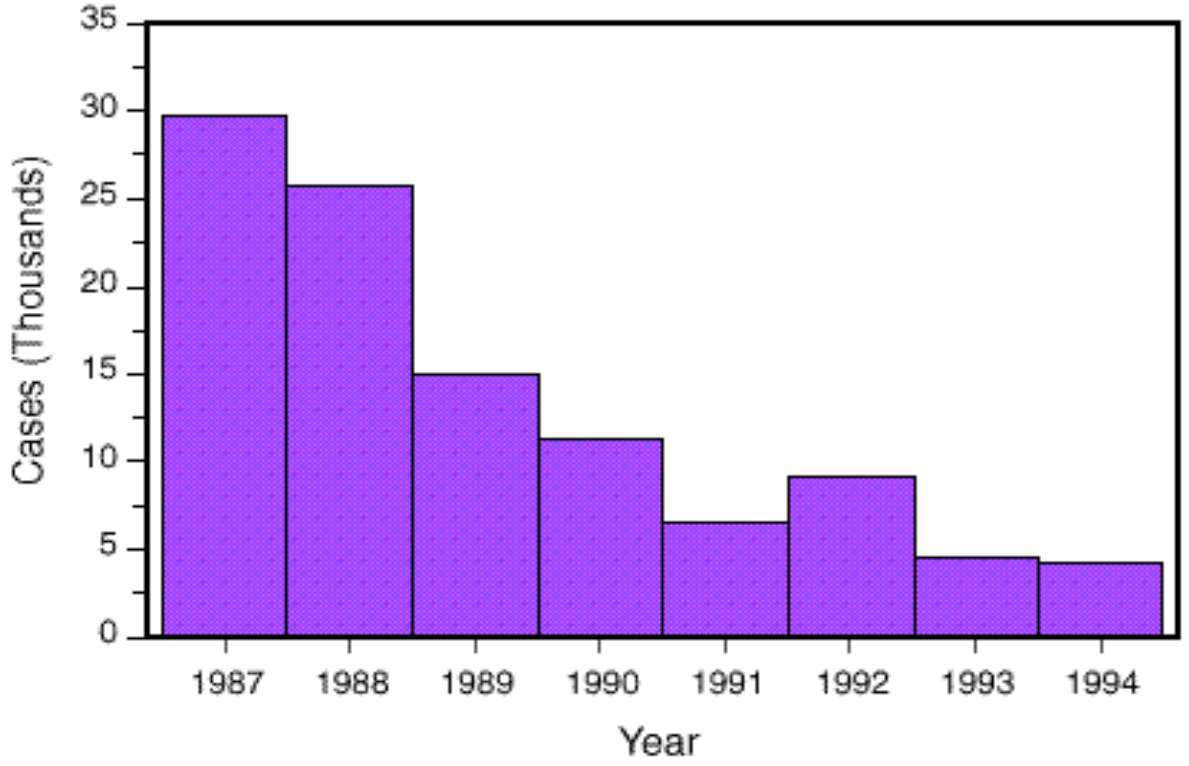


Figure 5–8 Survival curves over time, based on serum testosterone level, Eastern Cooperative Oncology Group.

122

FIGURE 1. Number of reported cases of paralytic poliomyelitis, by year — South East Asia Region (SEAR),* World Health Organization (WHO), 1987–1994



*Member countries of SEAR are Bangladesh, Bhutan, Democratic People's Republic of Korea (DPR Korea), India, Indonesia, Maldives, Myanmar, Nepal, Sri Lanka, and Thailand. Mongolia, formerly a member country, was administratively transferred in 1995 to the Western Pacific Region of WHO; data are included for Mongolia through 1994.

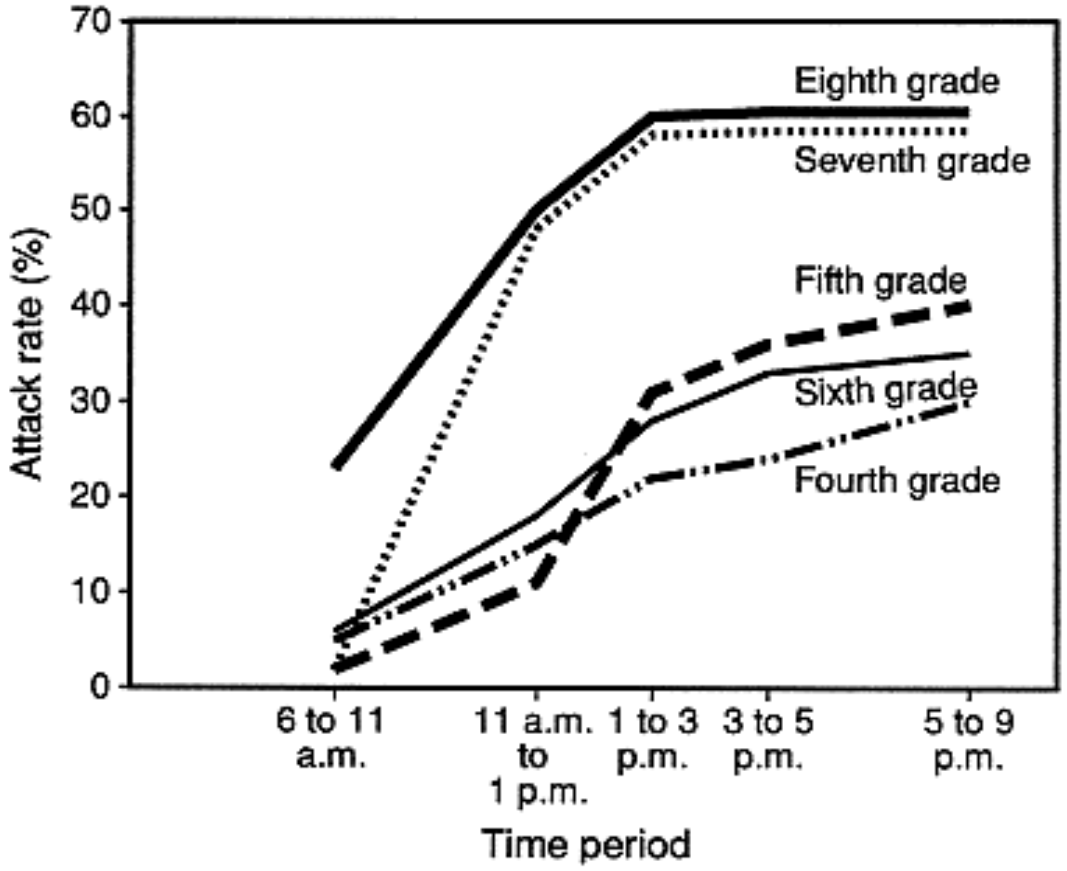


Figure 5–7 Sample cumulative attack rate, by grade in school and time of onset—North Carolina, 1985.

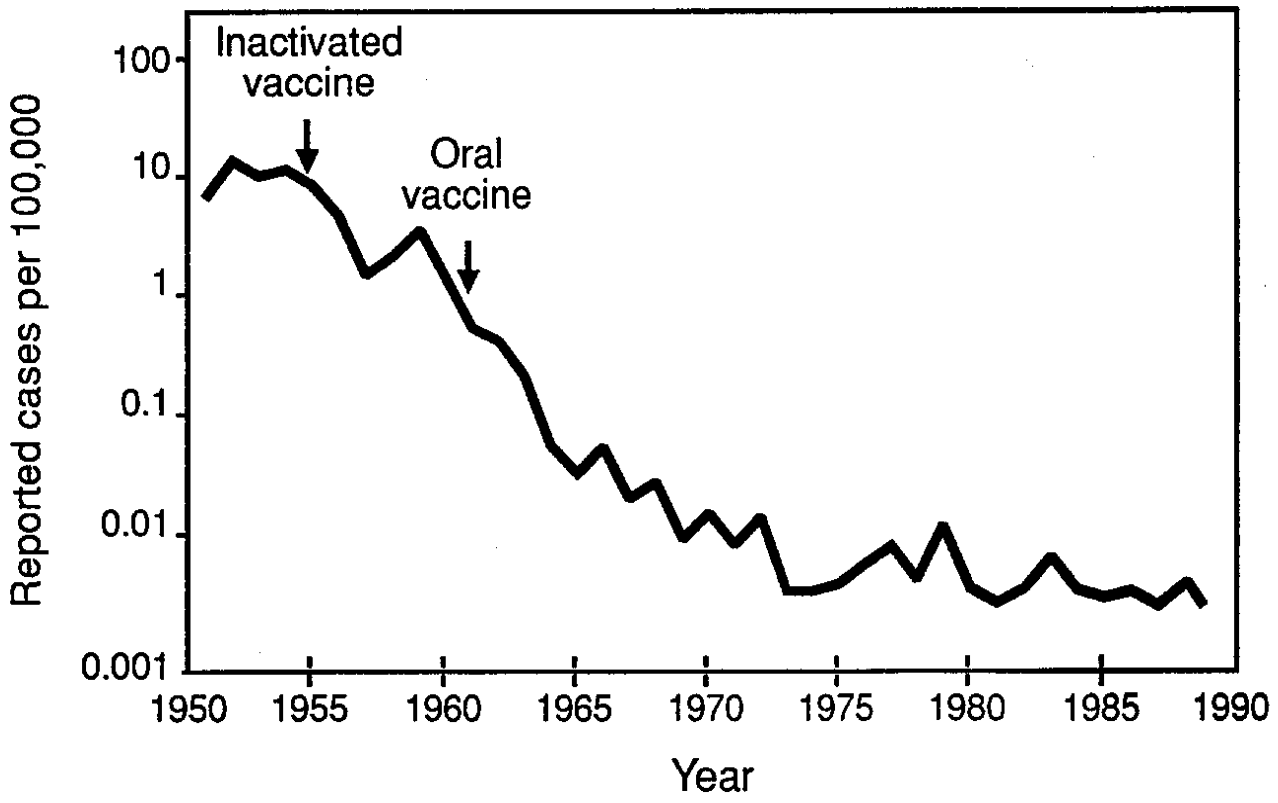


Figure 1–6 Semi-logarithmic-scale line graph of reported cases of paralytic poliomyelitis—United States, 1951–1989.

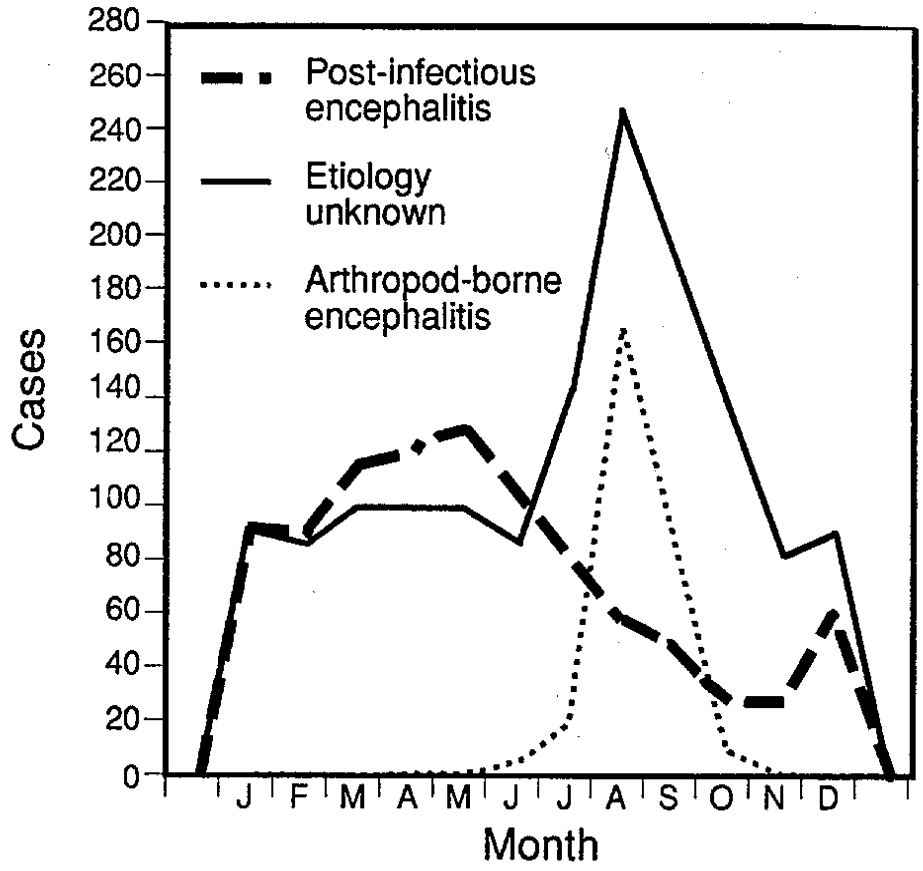
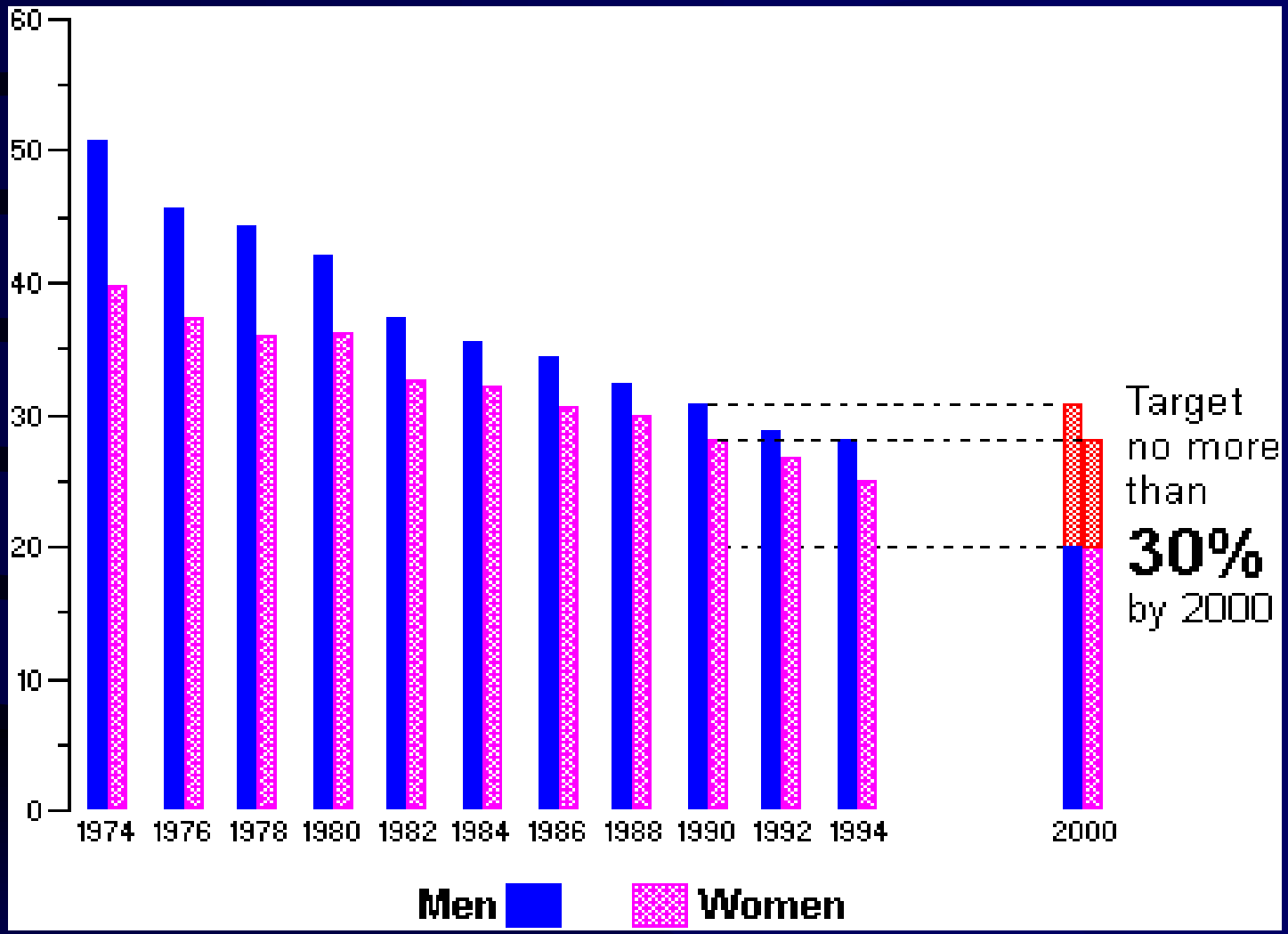


Figure 5–9 Frequency polygon of reported cases of encephalitis—United States, 1965.

England's population 1974-1994 and target for the year 2000 by sex, adults aged 16 and over



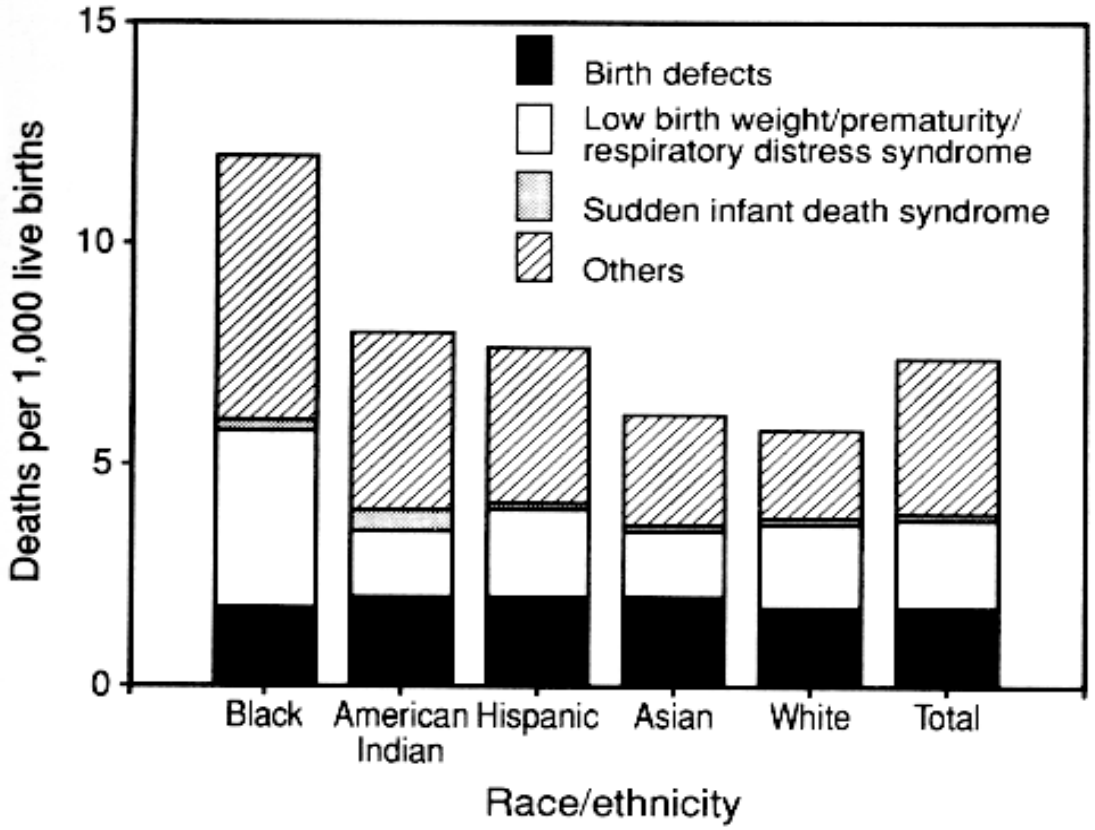
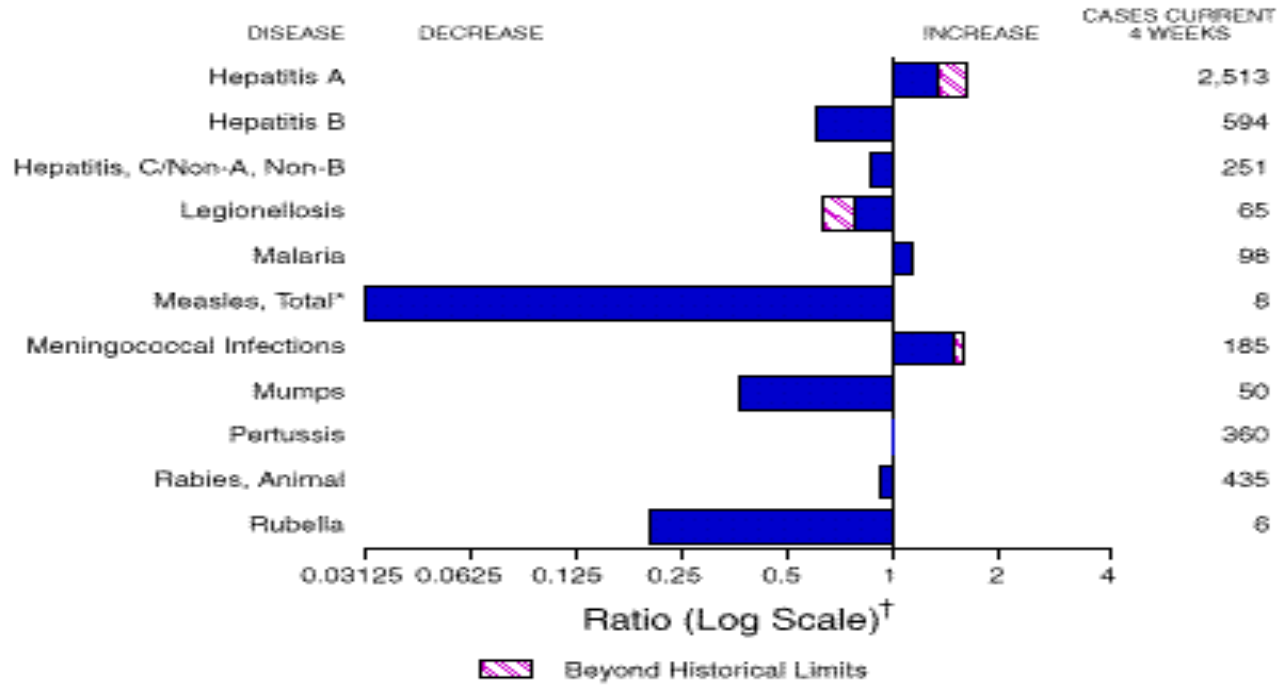


Figure 5–11 Stacked bar chart of underlying causes of infant mortality, by racial or ethnic group and age at death—United States, 1983.

FIGURE I. Notifiable disease reports, comparison of 4-week totals ending October 21, 1995, with historical data — United States



*The large apparent decrease in the number of reported cases of measles (total) reflects dramatic fluctuations in the historical baseline.

†Ratio of current 4-week total to mean of 15 4-week totals (from previous, comparable, and subsequent 4-week periods for the past 5 years). The point where the hatched area begins is based on the mean and two standard deviations of these 4-week totals.

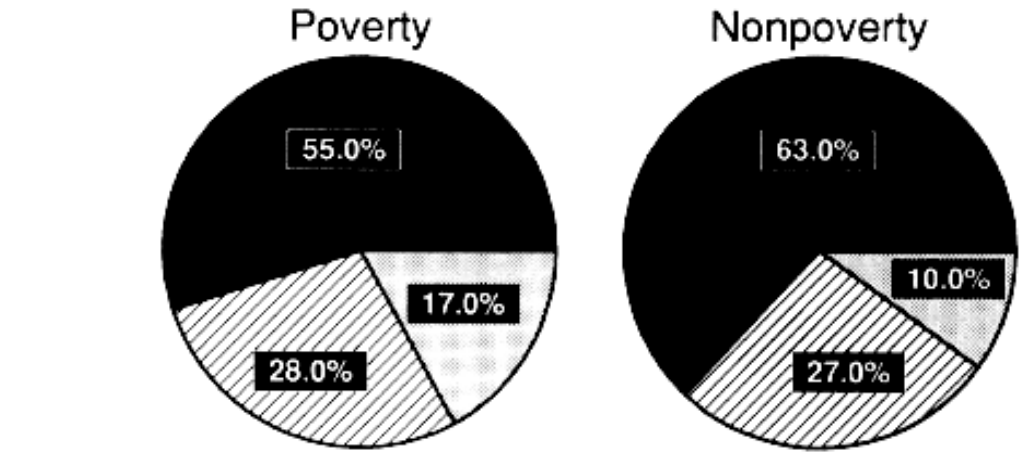
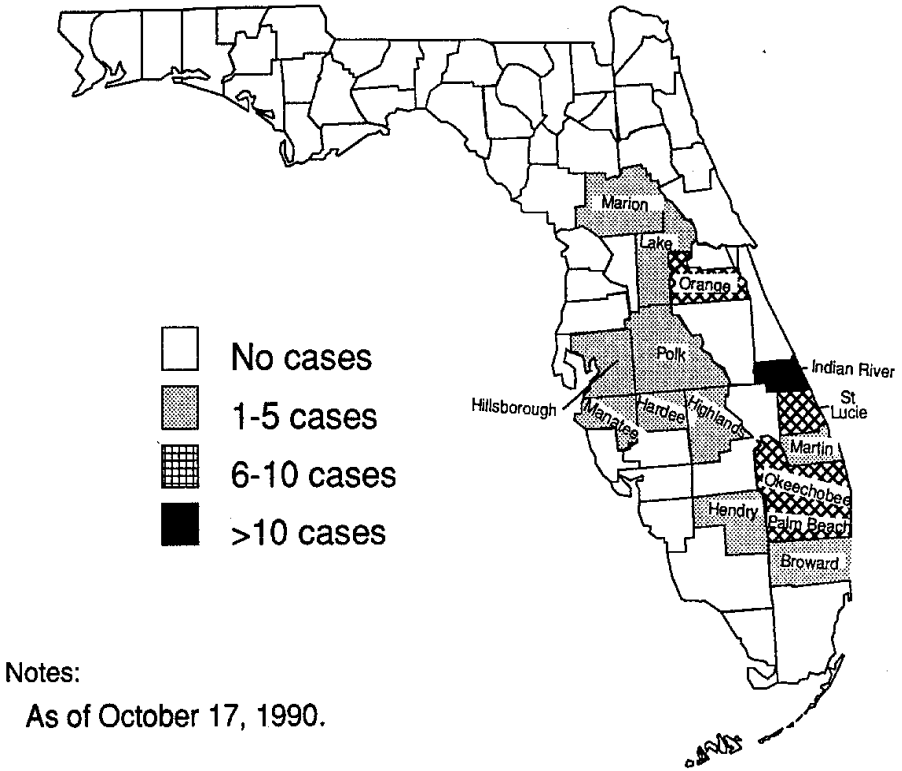


Figure 5–13 Pie charts of poliomyelitis vaccination status of children ages 1–4 years in cities with populations ≥ 250,000, by financial status—United States, 1969.

Maps

- Graphically represent data using location and geographic coordinates
- Provide a clear, concise, and quick method for grasping data

Figure 5–14 Spot map of deaths from smallpox—California, 1915–1924.



Notes:

As of October 17, 1990.

Figure 5–15 Chloropleth map of confirmed and presumptive cases of St. Louis encephalitis, by county—Florida, 1990.

Figure 5–16 Density-equalizing map of California (based upon population density), depicting deaths from smallpox, 1915–1924.

Limitations in Data

- Under-reporting
- Unrepresentativeness
- Inconsistent case definitions

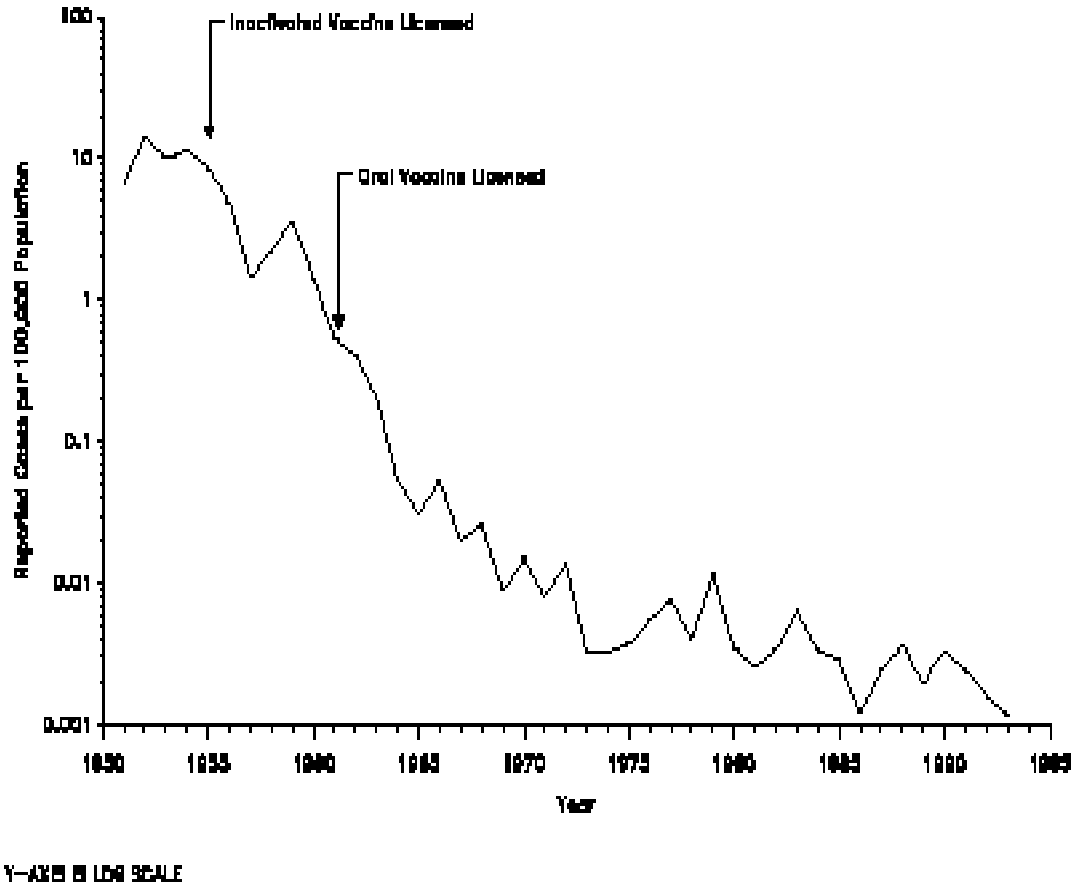
Considerations for Interpreting Data

- Has nature of reporting changed?
- Have new providers or geographic areas entered the system?
- Has case definition changed?
- Has new intervention been introduced?

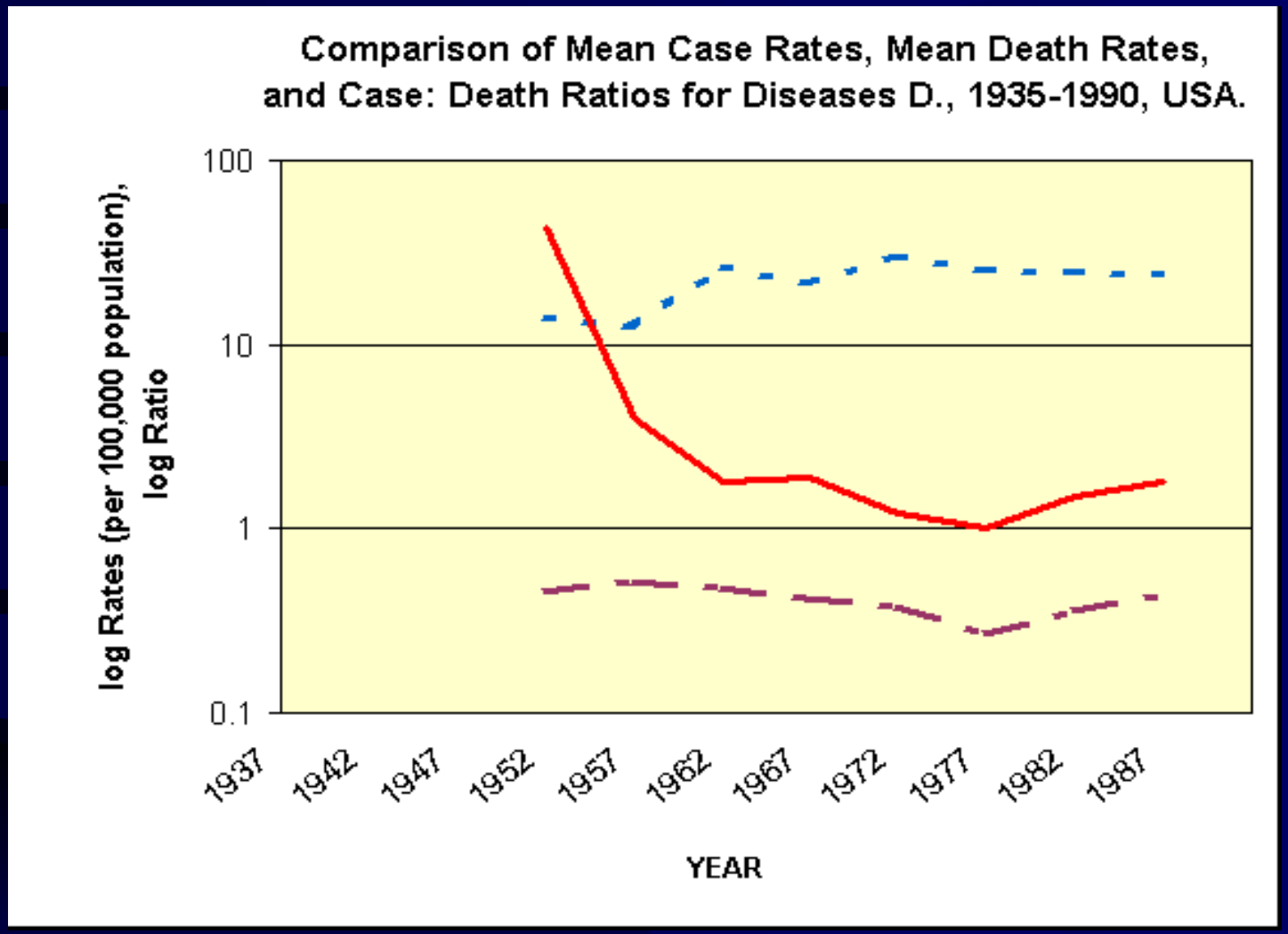
Interpretive Uses for Surveillance Data

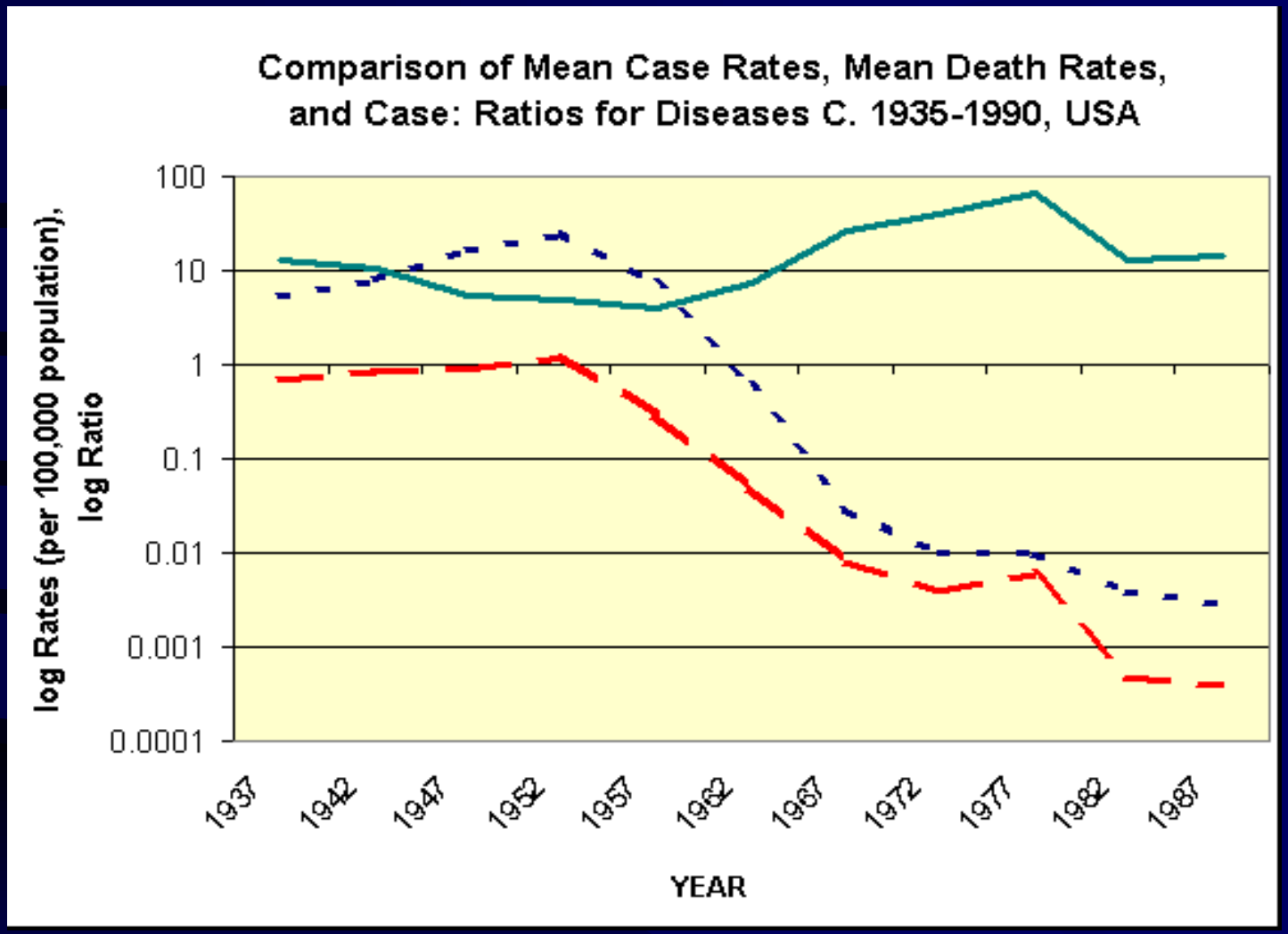
- Monitoring trends
- Identifying epidemics
- Identifying syndrome
- Evaluating public policy
- Projecting future needs

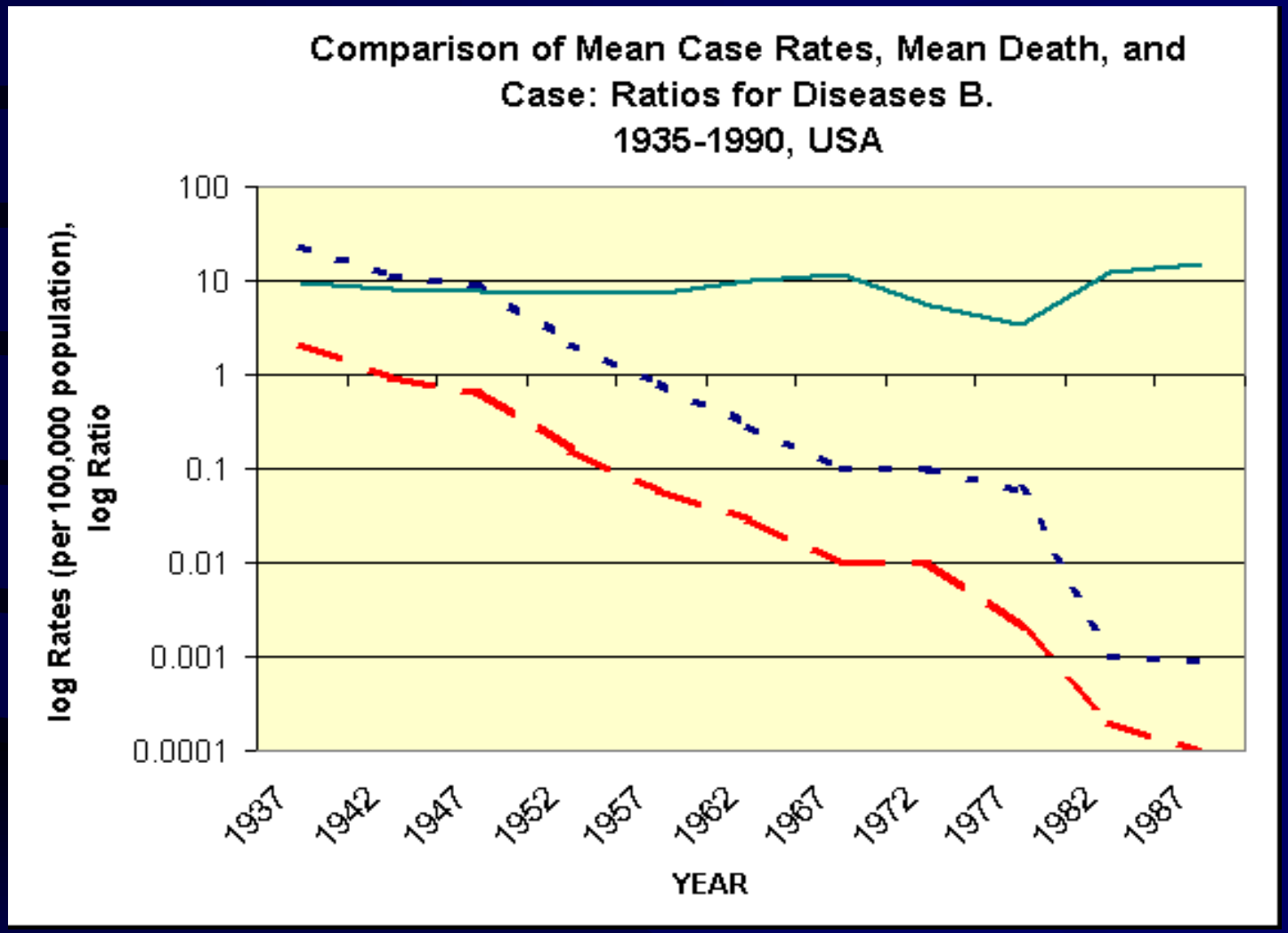
POLIOMYELITIS (paralytic) — by year, United States, 1951–1993

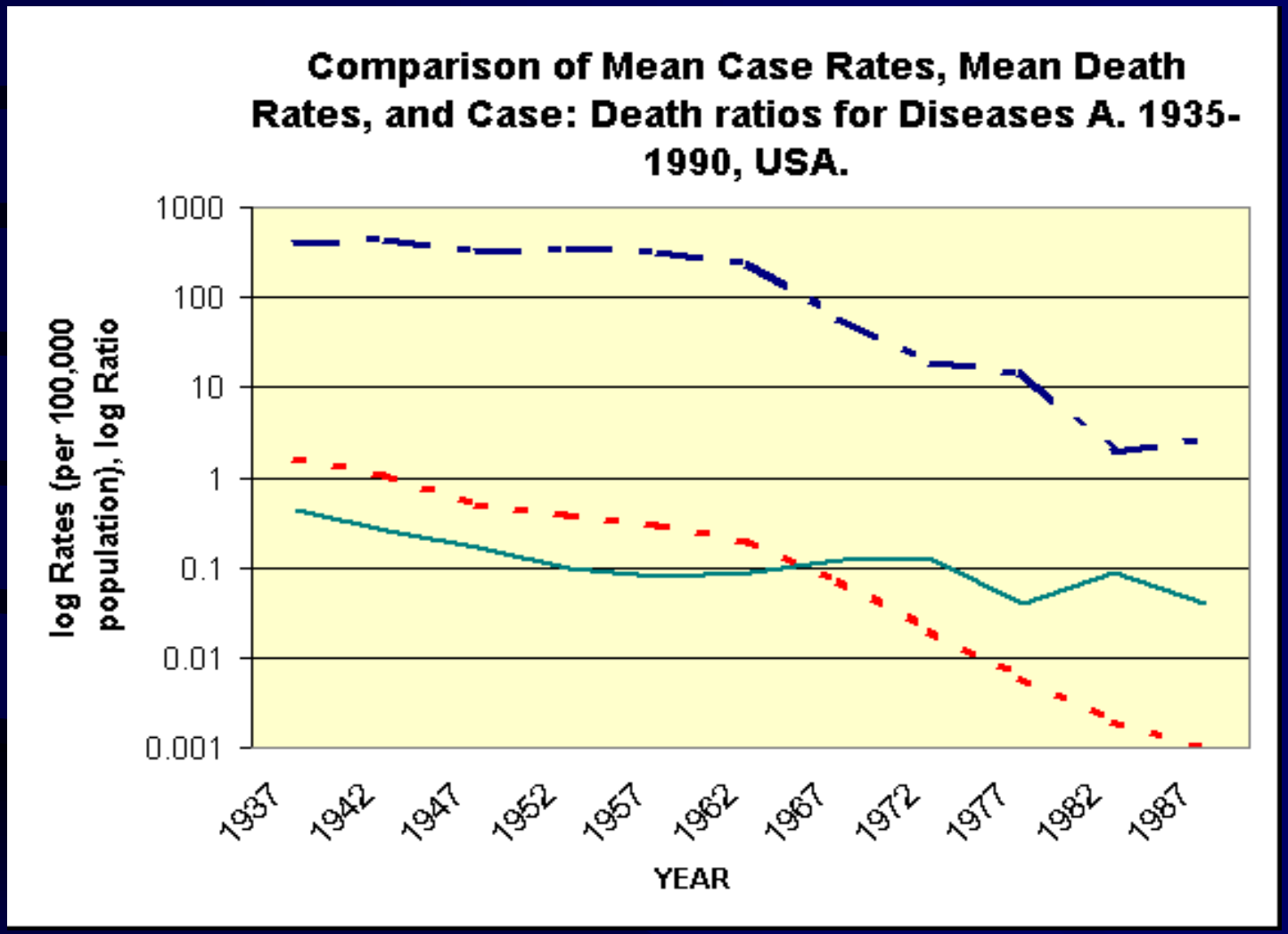


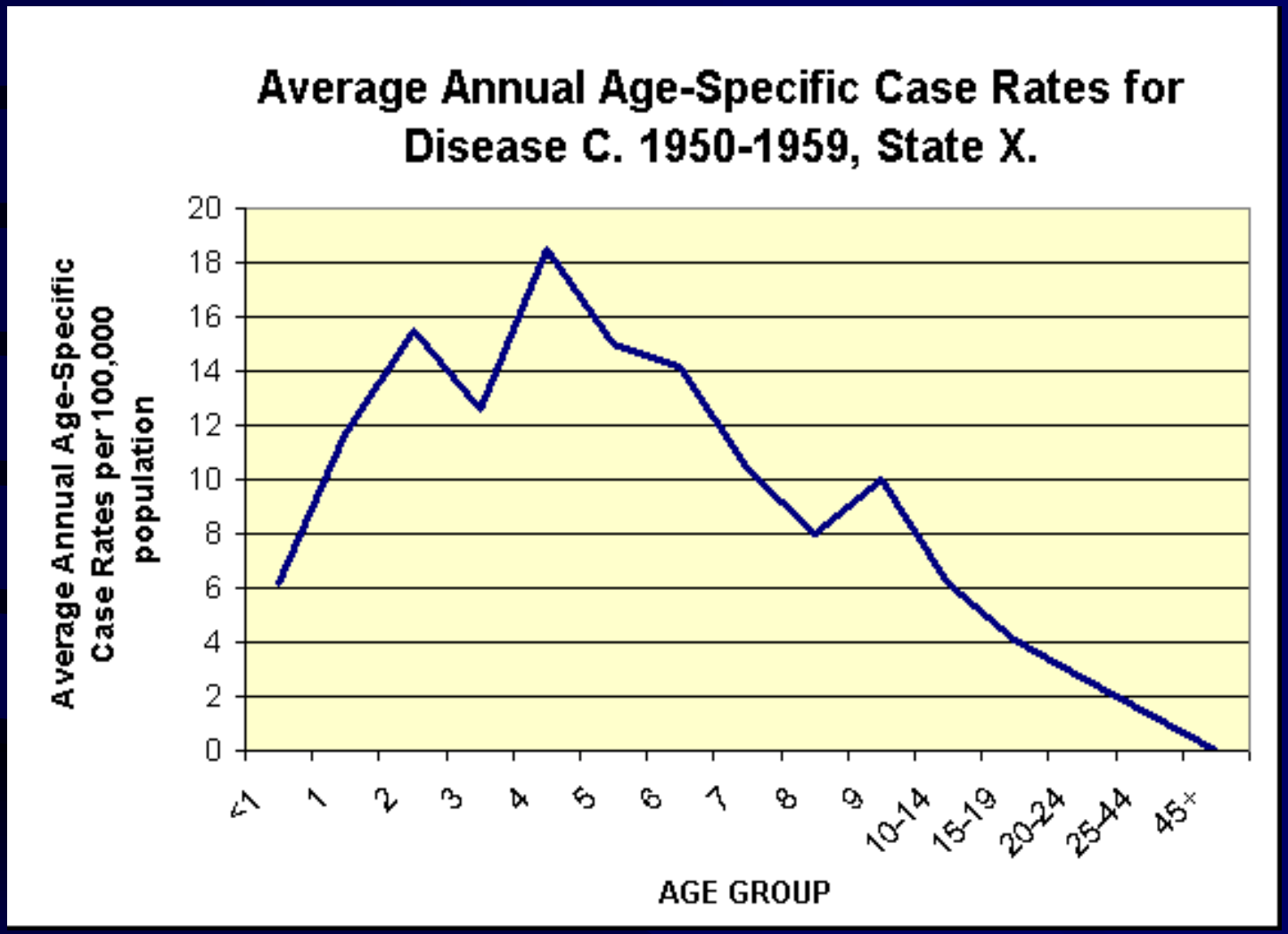
Y–AXIS IS LOG SCALE

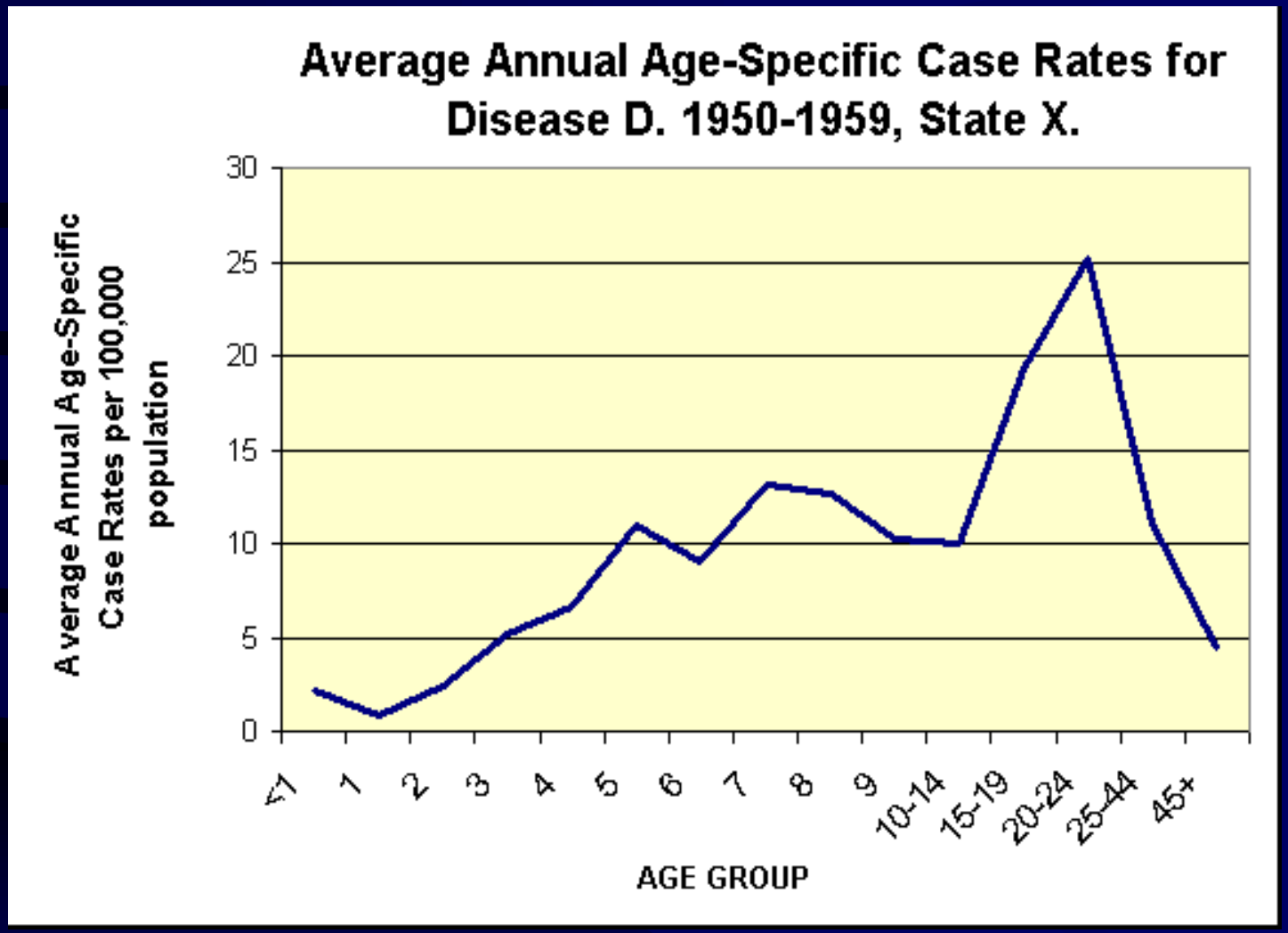

Comparison of Mean Case Rates, Mean Death Rates, and Case: Death Ratios for Diseases D., 1935-1990, USA.
log Rates (per 100,000 population), log Ratio
100
10
1
0.1
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR

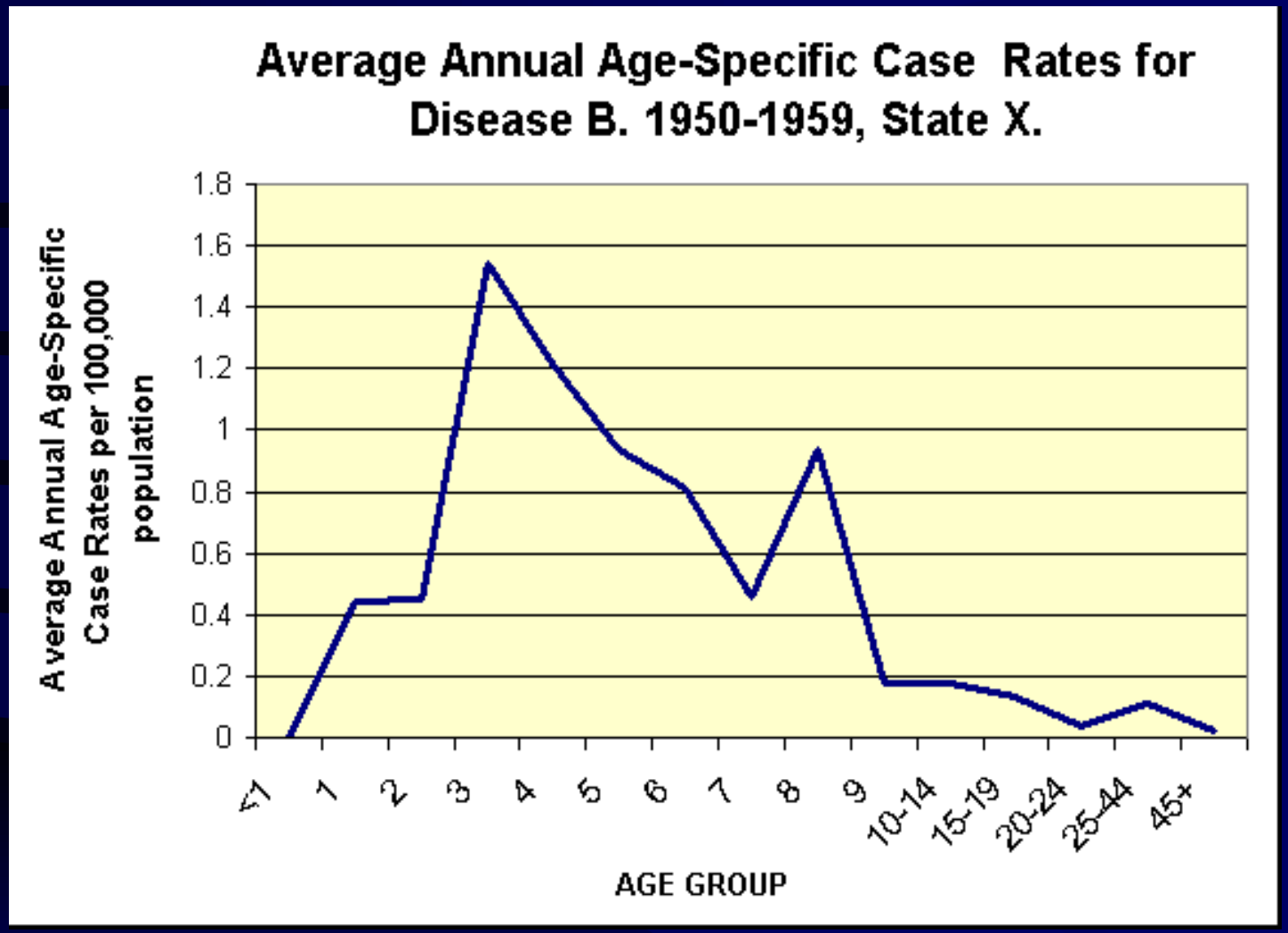

Comparison of Mean Case Rates, Mean Death Rates, and Case: Ratios for Diseases C. 1935-1990, USA
log Rates (per 100,000 population), log Ratio
100
10
1
0.1
0.01
0.001
0.0001
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR

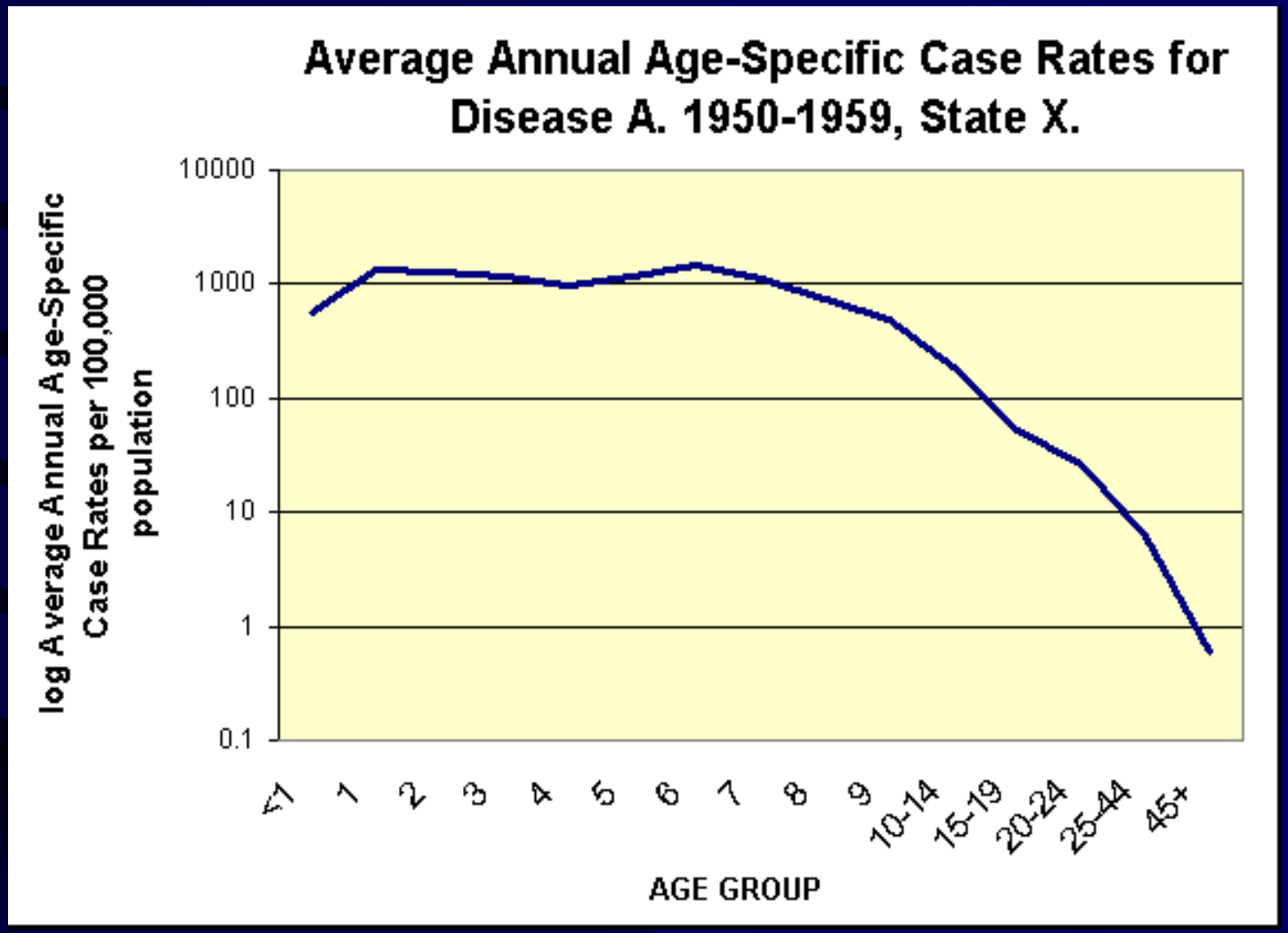

Comparison of Mean Case Rates, Mean Death, and Case: Ratios for Diseases B.
1935-1990, USA
log Rates (per 100,000 population), log Ratio
100
10
1
0.1
0.01
0.001
0.0001
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR

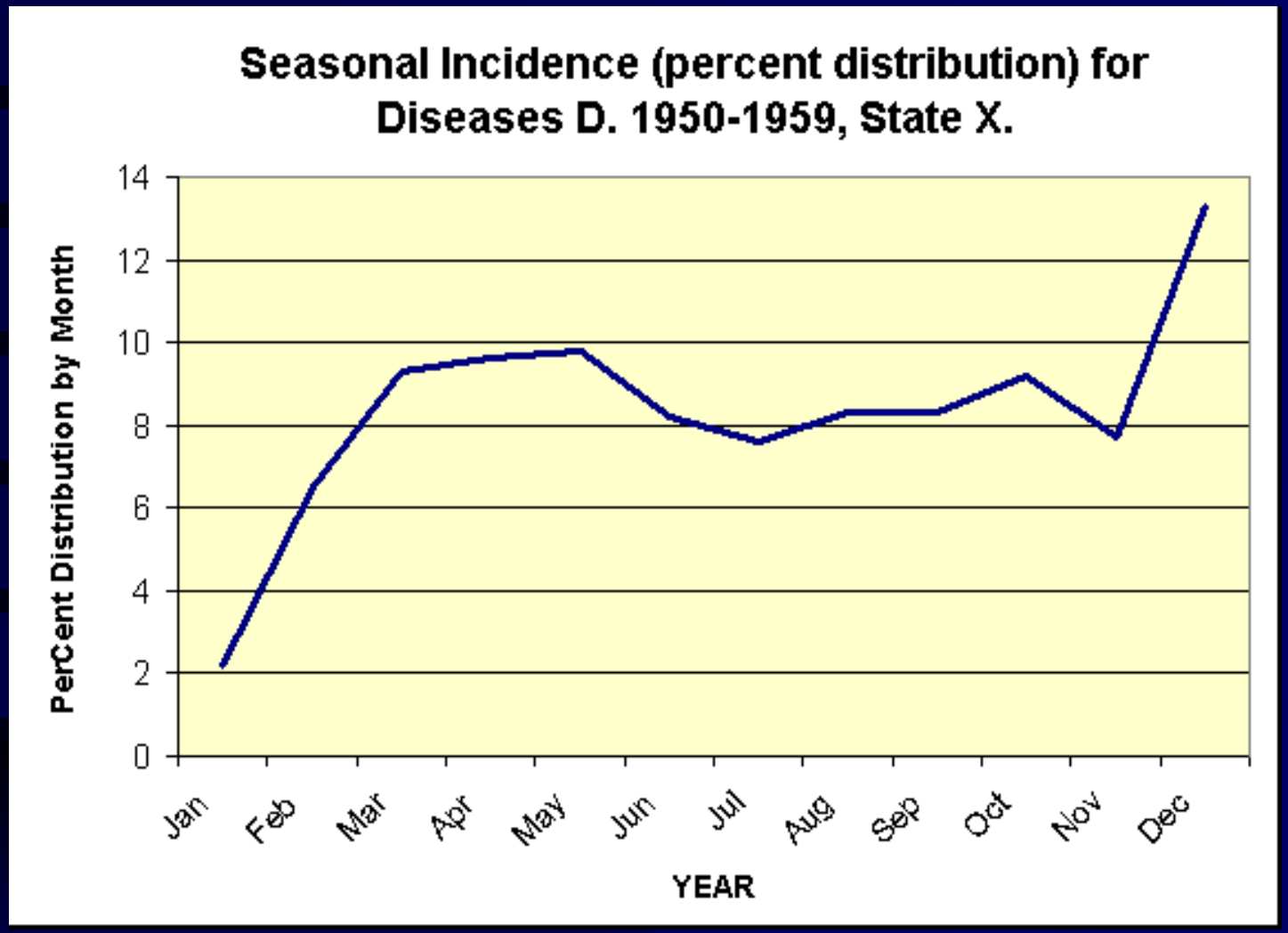

Comparison of Mean Case Rates, Mean Death Rates, and Case: Death ratios for Diseases A. 1935-1990, USA.
log Rates (per 100,000 population), log Ratio
1000
100
10
1
0.1
0.01
0.001
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR

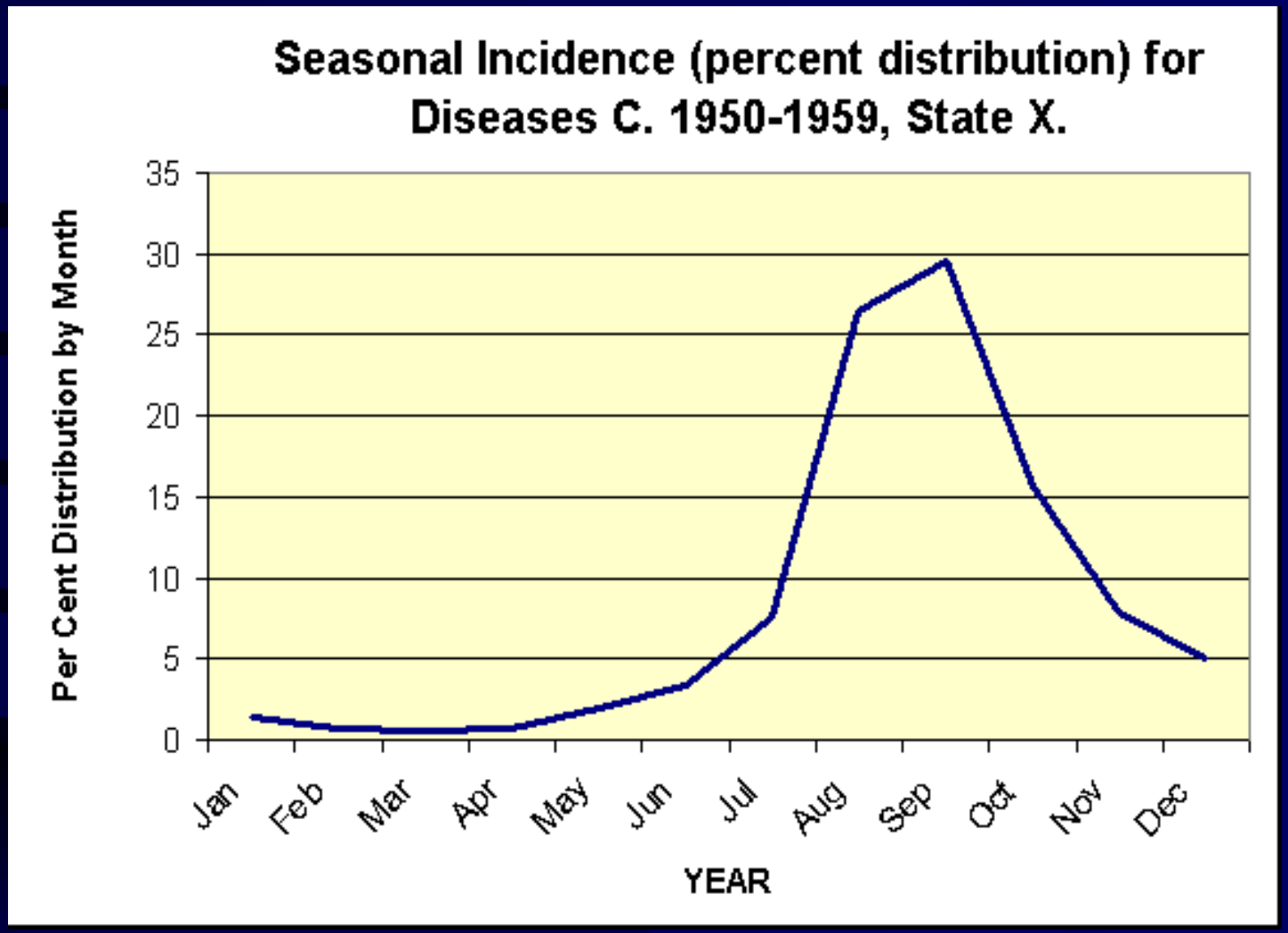

Average Annual Age-Specific Case Rates for Disease C. 1950-1959, State X.
Average Annual Age-Specific Case Rates per 100,000 population
20
18
16
14
12
10
8
6
4
2
0
<1
1
2
3
4
5
6
7
8
9
10-14
15-19
20-24
25-44
45+
AGE GROUP

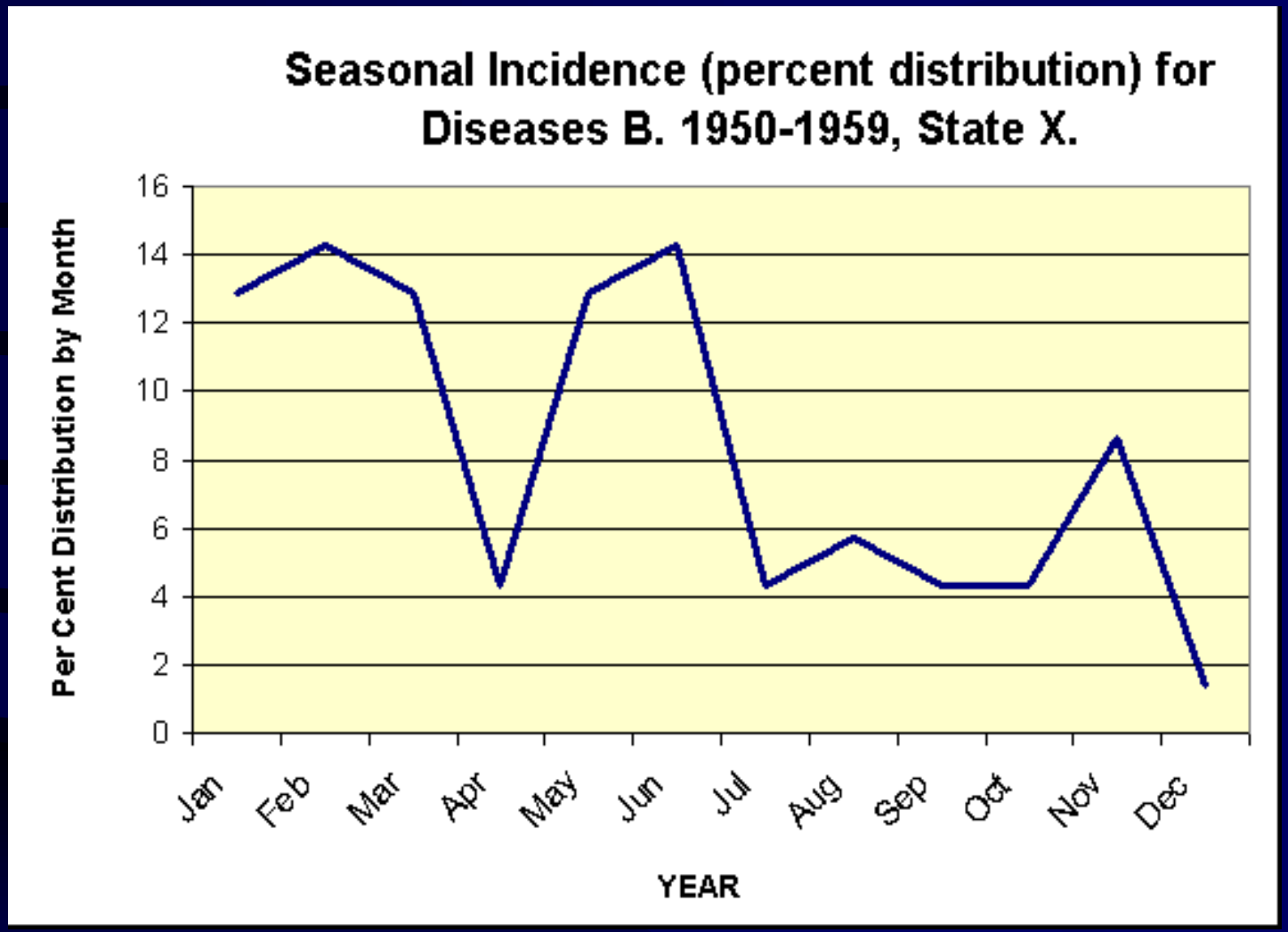

Average Annual Age-Specific Case Rates for Disease D. 1950-1959, State X.
Average Annual Age-Specific Case Rates per 100,000 population
30
25
20
15
10
5
0
<1
1
2
3
4
5
6
7
8
9
10-14
15-19
20-24
25-44
45+
AGE GROUP

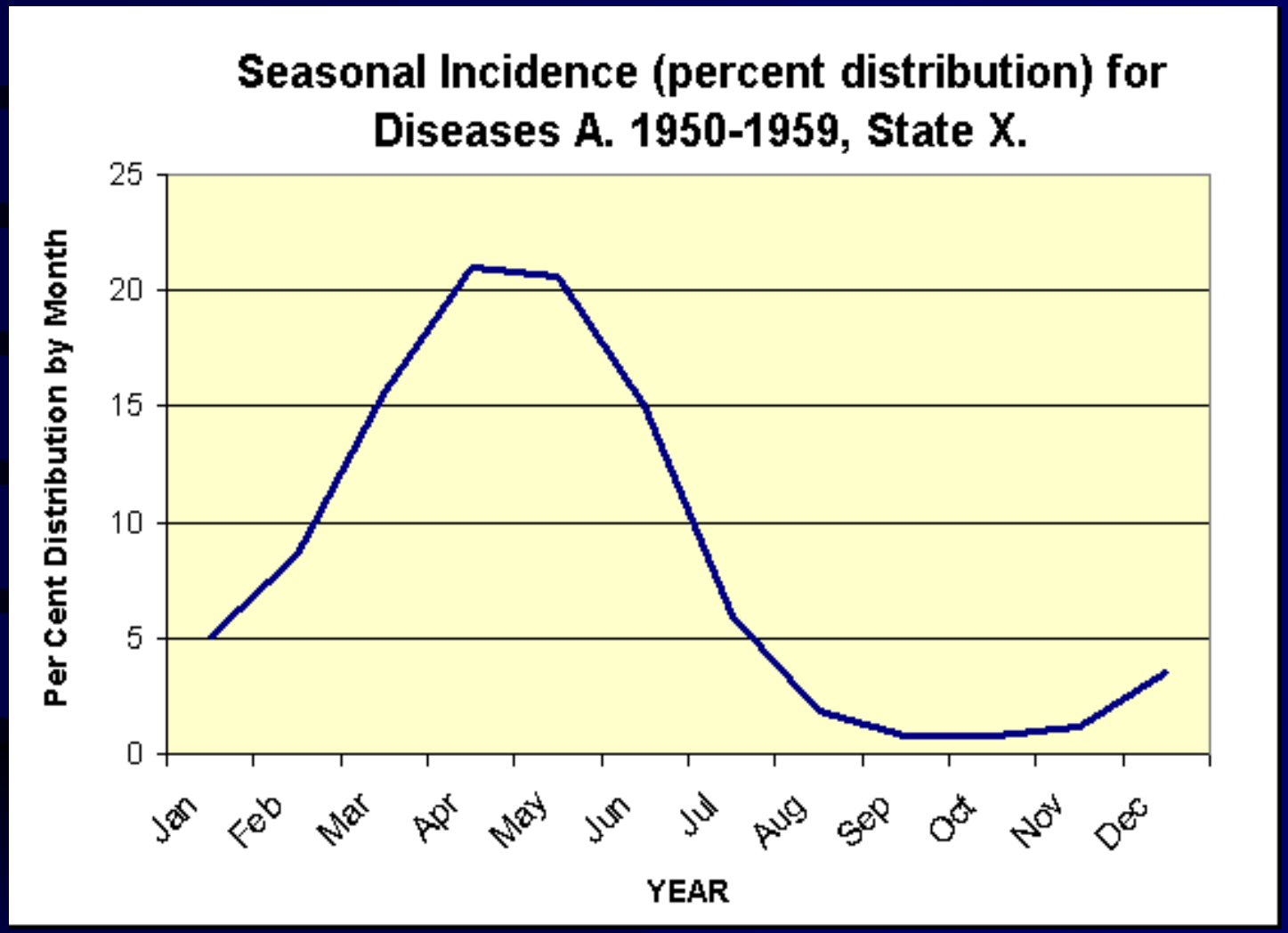

Average Annual Age-Specific Case Rates for Disease B. 1950-1959, State X.
Average Annual Age-Specific Case Rates per 100,000 population
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
<1
1
2
3
4
5
6
7
8
9
10-14
15-19
20-24
25-44
45+
AGE GROUP

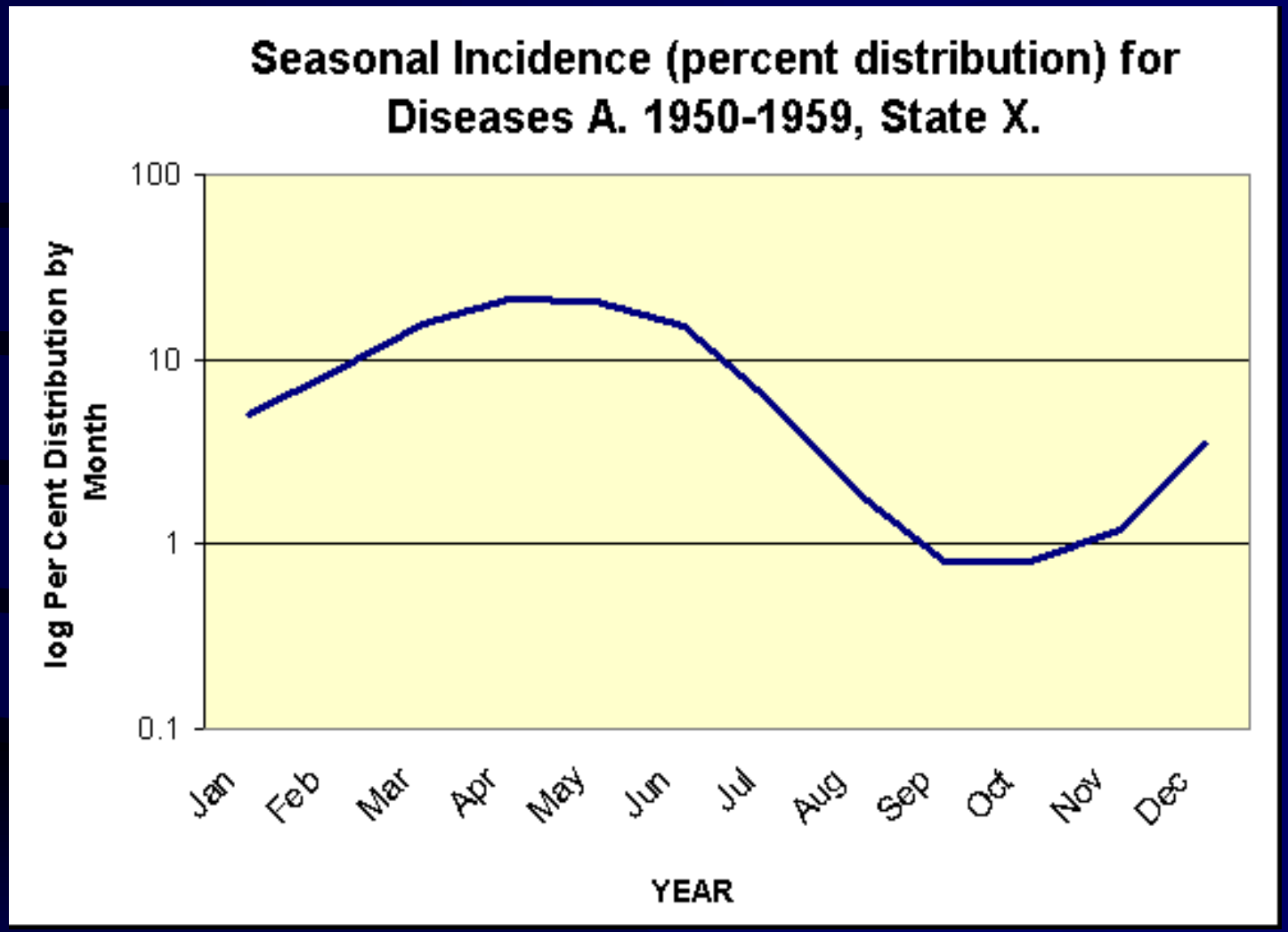

Average Annual Age-Specific Case Rates for Disease A. 1950-1959, State X.
log Average Annual Age-Specific Case Rates per 100,000 population
10000
1000
100
10
1
0.1
<1
1
2
3
4
5
6
7
8
9
10-14
15-19
20-24
25-44
45+
AGE GROUP

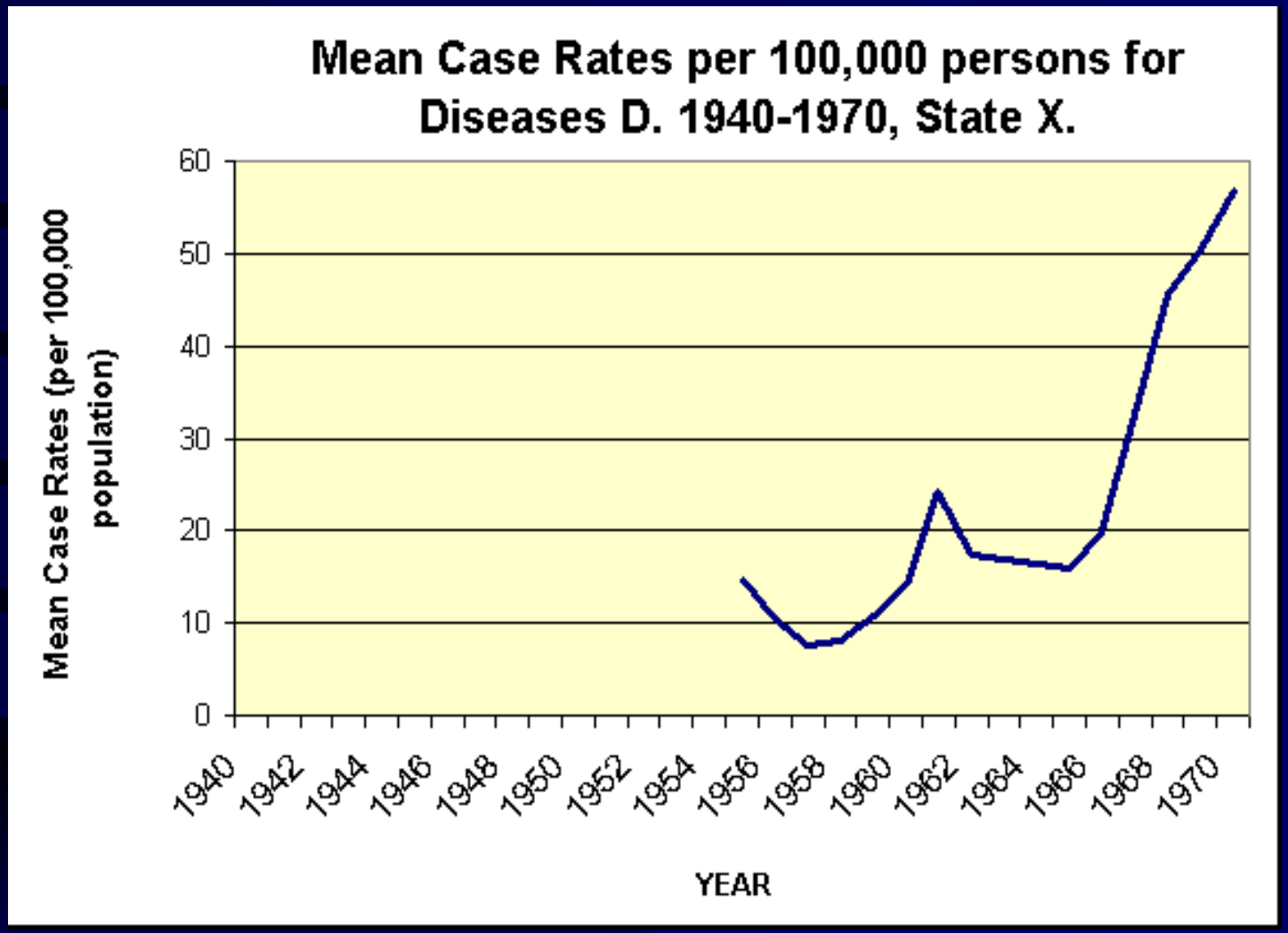

Seasonal Incidence (percent distribution) for Diseases D. 1950-1959, State X.
PerCent Distribution by Month
14
12
10
8
6
4
2
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
YEAR

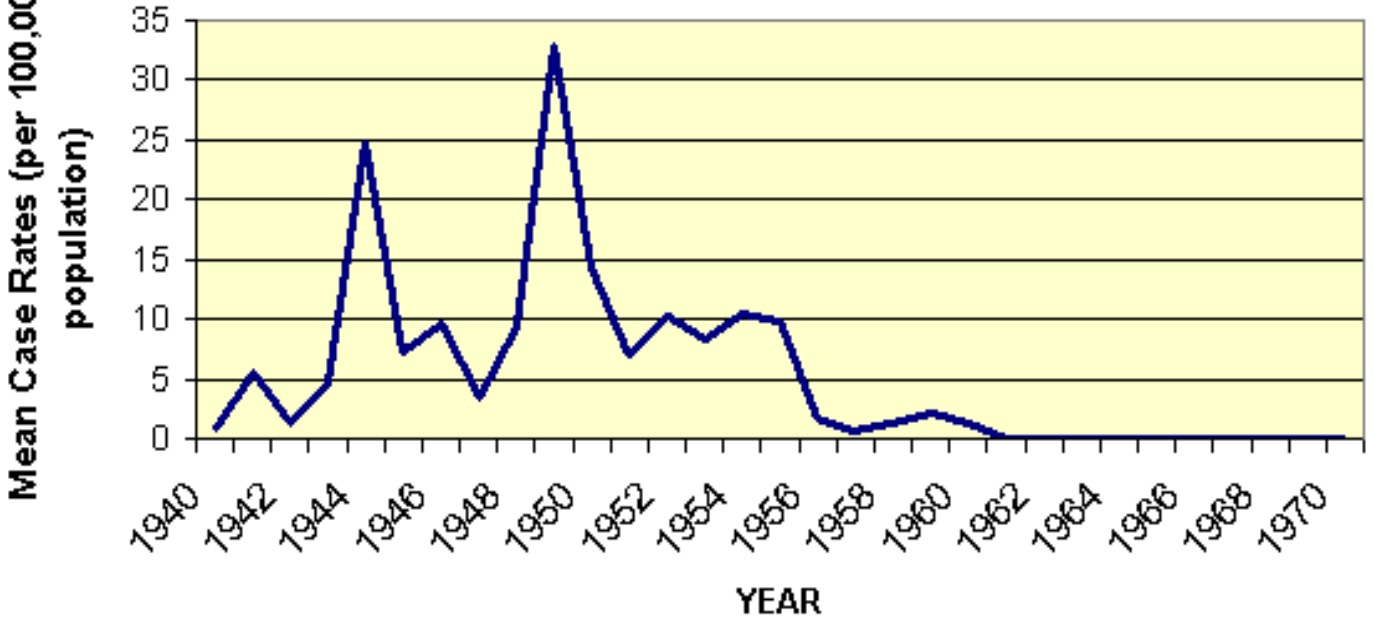

Seasonal Incidence (percent distribution) for Diseases C. 1950-1959, State X.
Per Cent Distribution by Month
35
30
25
20
15
10
5
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
YEAR

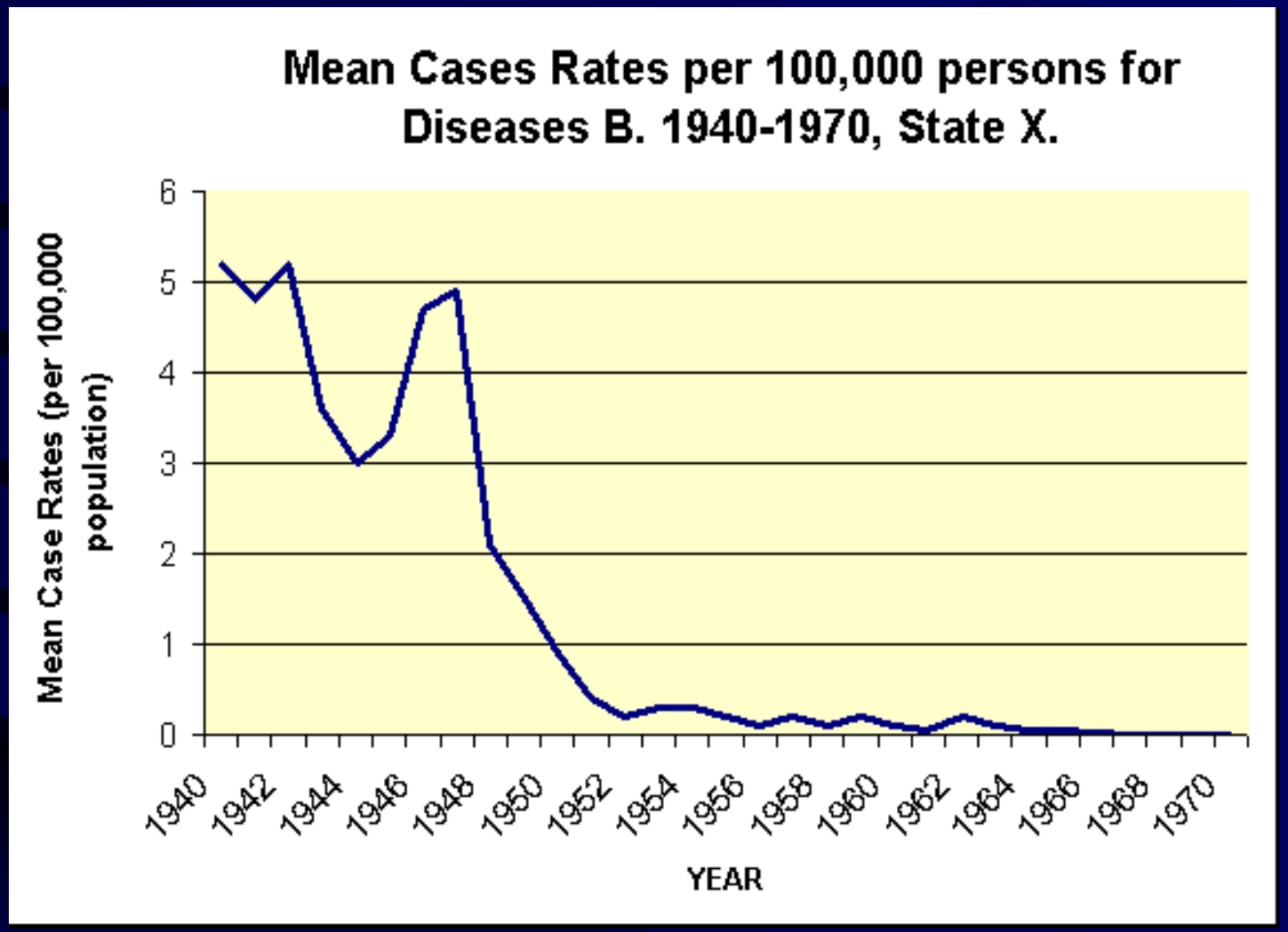

Seasonal Incidence (percent distribution) for Diseases B. 1950-1959, State X.
Per Cent Distribution by Month
16
14
12
10
8
6
4
2
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
YEAR

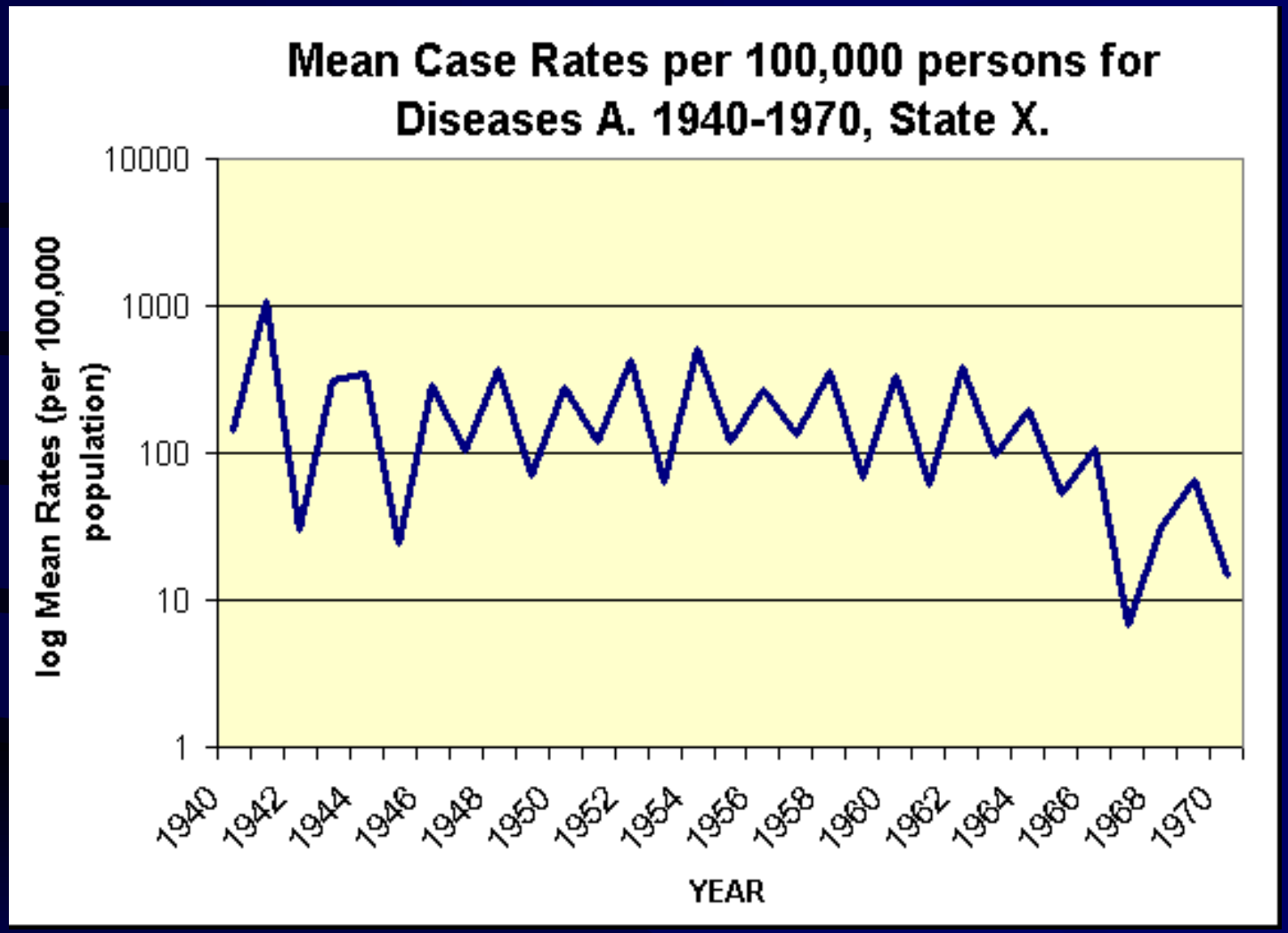

Seasonal Incidence (percent distribution) for Diseases A. 1950-1959, State X.
Per Cent Distribution by Month
25
20
15
10
5
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
YEAR

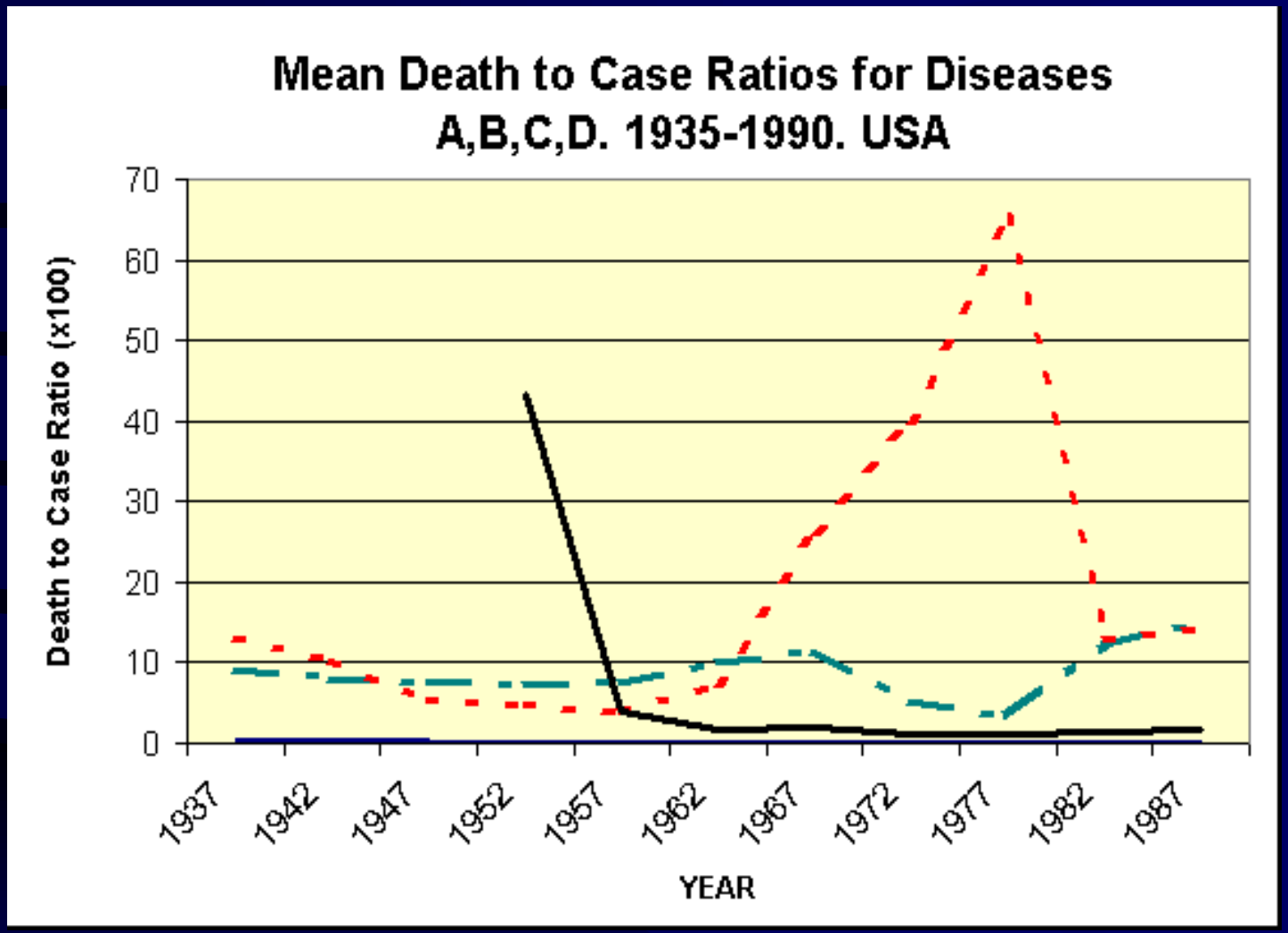

Seasonal Incidence (percent distribution) for Diseases A. 1950-1959, State X.
log Per Cent Distribution by Month
100
10
1
0.1
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
YEAR

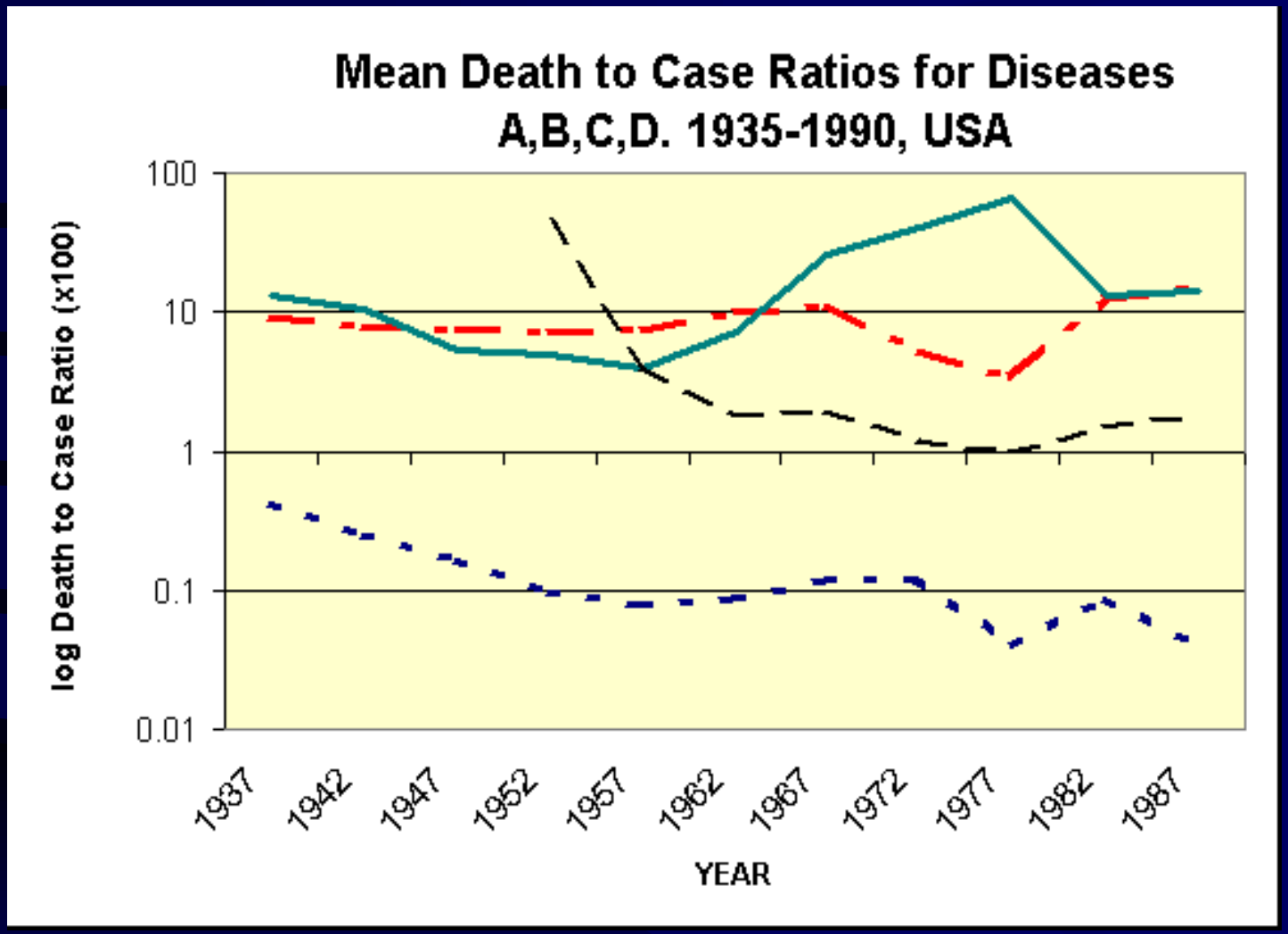

Mean Case Rates per 100,000 persons for Diseases D. 1940-1970, State X.
Mean Case Rates (per 100,000 population)
60
50
40
30
20
10
0
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
YEAR

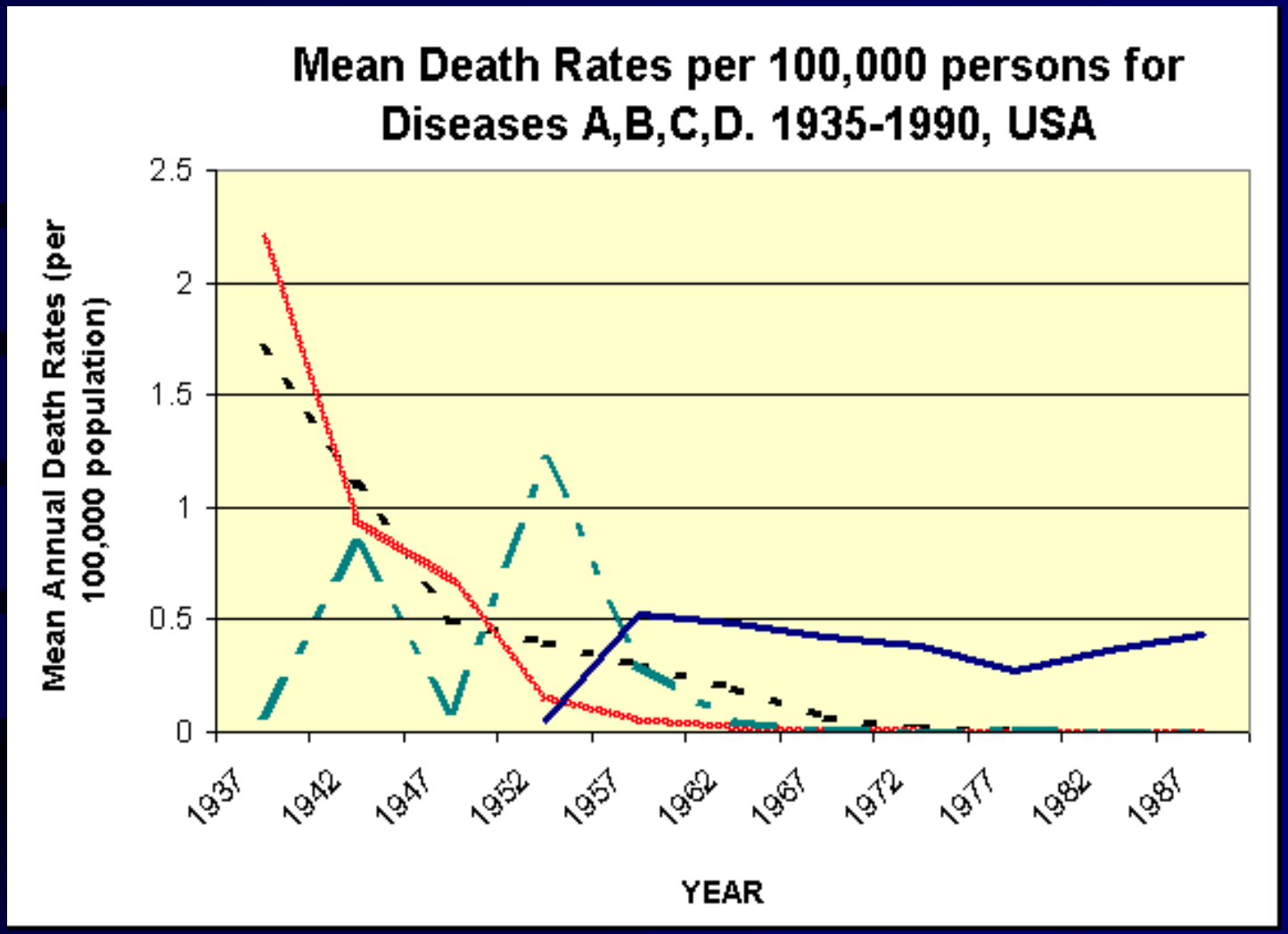

Mean Case Rates per 100,000 persons for Diseases C. 1940-1970, State X.
Mean Case Rates (per 100,000 population)
35
30
25
20
15
10
5
0
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
YEAR

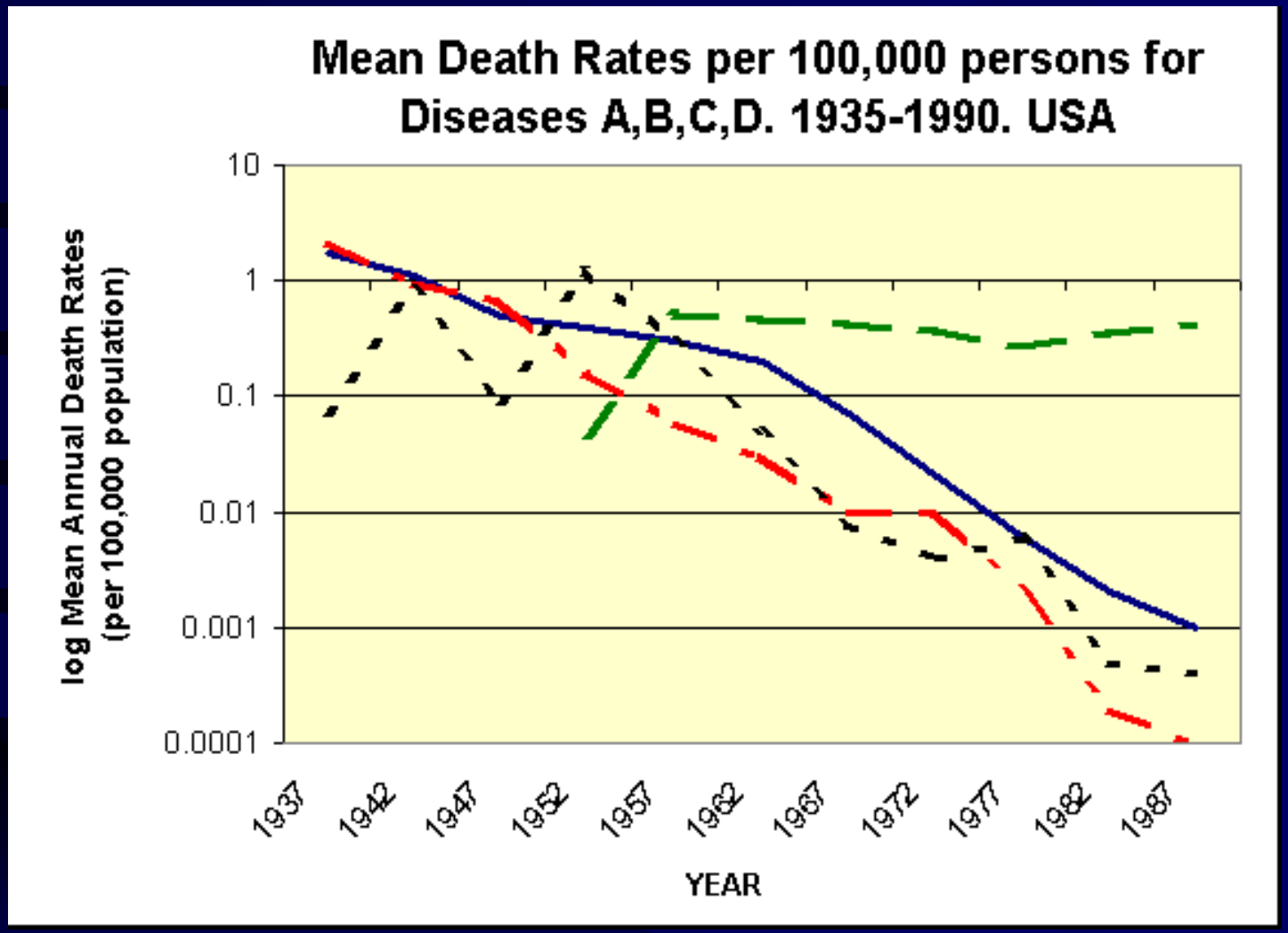

Mean Cases Rates per 100,000 persons for Diseases B. 1940-1970, State X.
Mean Case Rates (per 100,000 population)
0
1
2
3
4
5
6
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
YEAR

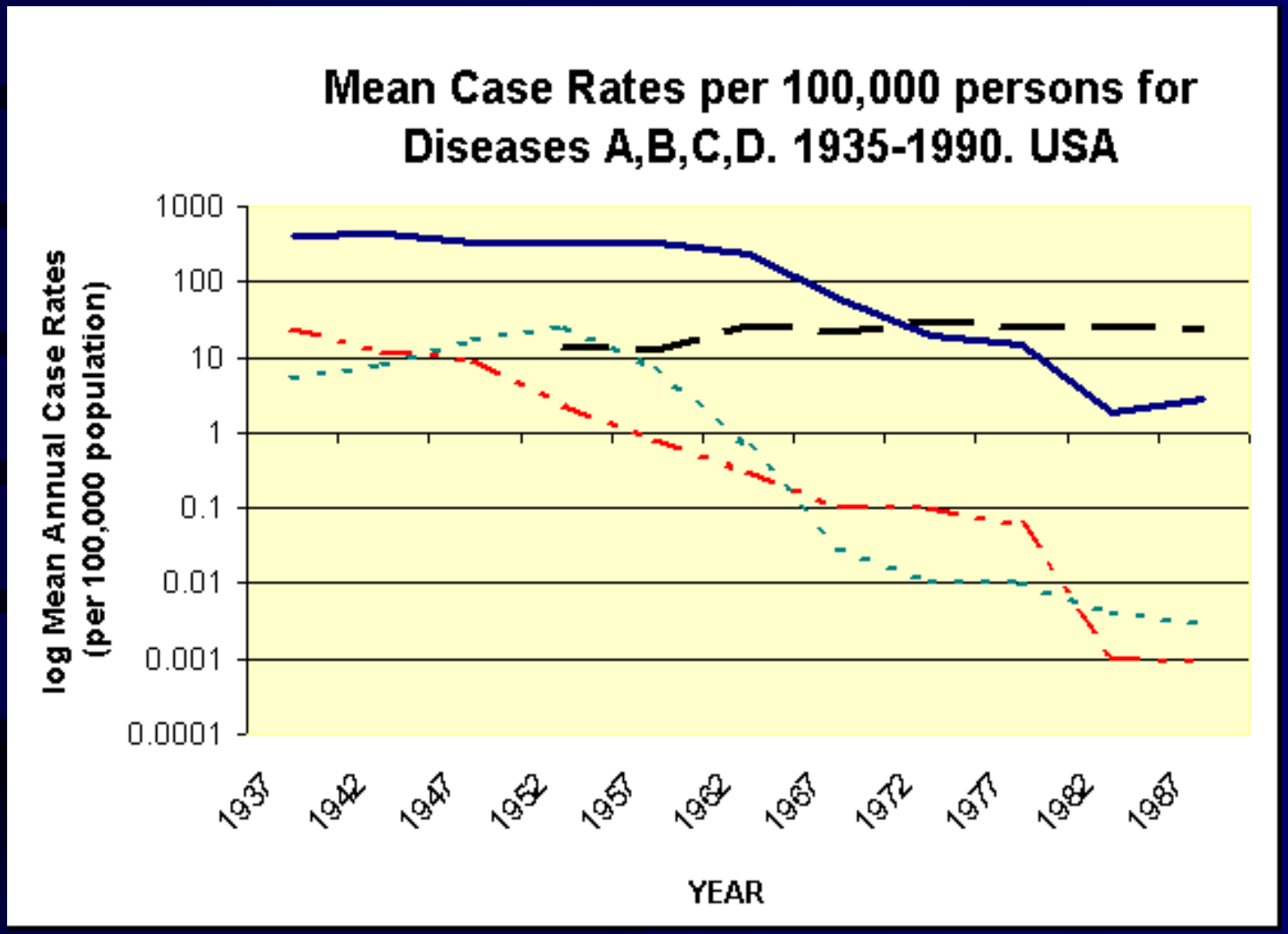

Mean Case Rates per 100,000 persons for Diseases A. 1940-1970, State X.
log Mean Rates (per 100,000 population)
10000
1000
100
10
1
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
YEAR


Mean Death to Case Ratios for Diseases
A,B,C,D. 1935-1990. USA
Death to Case Ratio (x100)
70
60
50
40
30
20
10
0
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR


Mean Death to Case Ratios for Diseases
A,B,C,D. 1935-1990, USA
log Death to Case Ratio (x100)
100
10
1
0.1
0.01
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR


Mean Death Rates per 100,000 persons for Diseases A,B,C,D. 1935-1990, USA
Mean Annual Death Rates (per 100,000 population)
2.5
2
1.5
1
0.5
0
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR


Mean Death Rates per 100,000 persons for Diseases A,B,C,D. 1935-1990. USA
log Mean Annual Death Rates (per 100,000 population)
10
1
0.1
0.01
0.001
0.0001
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR


Mean Case Rates per 100,000 persons for Diseases A,B,C,D. 1935-1990. USA
log Mean Annual Case Rates (per 100,000 population)
1000
100
10
1
0.1
0.01
0.001
0.0001
1937
1942
1947
1952
1957
1962
1967
1972
1977
1982
1987
YEAR